Principles of Gamma Synchrony Predict Figure–Ground Perception in Texture Stimuli

Maryam Karimian[1,2,3,4], Mark J. Roberts[1,5], Peter De Weerd[1,2,5†], and Mario Senden[1,5†*]

[1]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

[2]Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, 6229 ER Maastricht, The Netherlands

[3]Institute for Theoretical Biology, Department of Biology, Humboldt-Universität zu Berlin, 10115 Berlin, Germany

[4]Science of Intelligence, Research Cluster of Excellence, 10587 Berlin, Germany

[5]Maastricht Brain Imaging Centre, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

† These authors contributed equally to this work.

* Corresponding author: Mario Senden (mario.senden@maastrichtuniversity.nl)

Abstract

Gamma synchrony is ubiquitous in visual cortex, but whether it contributes to perceptual grouping remains contentious based on observations that gamma frequency is not consistent across stimulus features and that gamma synchrony depends on distances between image elements. These stimulus dependencies have been argued to challenge the idea that the visual system groups image elements by synchronizing the neural assemblies that encode them. Here we argue instead that these dependencies may shape synchrony in perceptually meaningful ways. Indeed, according to the theory of weakly coupled oscillators (TWCO), synchrony-based grouping mechanisms require stimulus dependence. Synchronization among coupled oscillators depends on frequency dissimilarity and coupling strength, which in early visual cortex relate to local feature dissimilarity and physical distance, respectively. We manipulated these factors in a texture segregation experiment wherein human observers identified the orientation of a figure defined by reduced contrast heterogeneity compared to the background. Human performance followed TWCO predictions both qualitatively and quantitatively, as formalized in a computational model. Moreover, we found that when enriched with a Hebbian learning rule, our model also predicted human learning effects: Increases in model gamma synchrony due to perceptual learning predicted improvements in texture segregation across sessions. Taken together, our data suggest that the stimulus-dependence of gamma synchrony captures local image statistics and is linked to the stimulus-dependence of texture segregation, and that the effect of visual experience on gamma synchrony provides a viable perceptual learning mechanism for training-induced improvements in texture segregation. Our results suggest that gamma synchrony with its inherent stimulus dependencies can provide a plausible mechanistic basis for perceptual grouping and visual scene segmentation.

Introduction

Oscillations are ubiquitous in the cortex (Buzsáki et al., 2013) and can synchronize both within and between cortical areas (Anand et al., 2023; Lowet, Roberts, Peter, Gips, & De Weerd, 2017; Melloni et al., 2007), but whether this contributes to neural information processing remains a matter of debate (Doelling & Florencia Assaneo, 2021; Duecker et al., 2021; Fernandez-Ruiz et al., 2023; Ray & Maunsell, 2015; Roelfsema, 2023). Early suggestions that synchrony in the gamma frequency band (30 – 80 Hz) plays a central role in visual feature binding (Singer, 1999; Uhlhaas et al., 2008) have been called into question based on observations that the gamma frequency is not consistent across stimulus features (Ray & Maunsell, 2010, 2015; Shirhatti et al., 2022) and depends on distances between image elements (Roelfsema, 2023; Roelfsema et al., 2004), making it difficult to group components of the same object by synchrony among the neural assemblies encoding these components (Dubey & Ray, 2020; Roelfsema, 2023). Alternatively, it has been proposed that the stimulus dependence of gamma synchrony facilitates, rather than hinders, their functional significance for visual processing by allowing contiguous neural assemblies that share a sufficiently similar oscillation frequency to synchronize into meaningful groups, while also blocking synchrony among assemblies with substantial frequency difference or physical separation (Lowet et al., 2015; Lowet, Roberts, Peter, Gips, & De Weerd, 2017). Here we show empirical and computational support for this view.

Analyzing a visual scene requires integration of features into coherent objects (feature binding), but also segregation of features belonging to distinct objects (feature separation). It remains unclear how this is achieved, but the stimulus dependence of gamma may be critical for a synchrony-based neural grouping mechanism that achieves both feature binding and separation. This idea is rooted in the theory of weakly coupled oscillators (TWCO), which describes the preconditions for synchrony among coupled oscillators (Acebrón et al., 2005; Ermentrout et al., 2019; Kuramoto, 1984; Neu, 1979; Strogatz, 2000). A group of coupled oscillators synchronizes if the discrepancy in their frequencies, referred to as their detuning, is overcome by the strength of their connection, referred to as their coupling. Thus, synchrony can occur even in the presence of strong detuning, if the coupling strength is sufficiently high, whereas if the coupling strength is low, synchrony can only occur if the detuning is also minimal. This relationship can be graphically depicted in an Arnold tongue (Coombes & Bressloff, 1999; Pikovsky et al., 2001), which shows the regions where synchrony occurs based on the balance between detuning and coupling strength (see Figure 1a for an illustration). These abstract principles are concretely realized in early visual cortex. Neural assemblies exhibit gamma oscillations in their population activity at frequencies that are directly related to stimulus features such as spatial frequency, contrast and orientation (Dubey & Ray, 2020; Henrie & Shapley, 2005; Shapira et al., 2017), and particularly contrast (Hadjipapas et al., 2015; Lowet et al., 2015; Roberts, Lowet, Brunet, TerWal, Tiesinga, Fries, & DeWeerd, 2013). In early visual cortical areas, coupling strength between neural assemblies is directly related to the efficacy of lateral anatomical connectivity, which declines with cortical distance (Boucsein et al., 2011; Gilbert & Wiesel, 1983; Lowet et al., 2015; Lowet, Roberts, Peter, Gips, & De Weerd, 2017; Stettler et al., 2002; Ts'o et al., 1986). In conjunction with the retinotopic organization of early visual cortex, this implies that neural assemblies encoding nearby visual regions are more strongly coupled. Taken together, synchrony in early visual cortex could occur across widely spaced neuronal assemblies in response to scenes with low feature heterogeneity, but only for closely spaced assemblies in response to scenes with high feature heterogeneity (see Figure 1b for an illustration). Indeed, a recent electrophysiological study in macaque V1 in which cortical distance and stimulus contrast heterogeneity were parametrically manipulated has confirmed that gamma synchrony behaves in line with the principles of TWCO (Lowet, Roberts, Peter, Gips, & De Weerd, 2017)

A synchrony-based grouping mechanism based on these principles has been successfully exploited for image segmentation in machine vision (Fang et al., 2014; Lowet et al., 2015; Nikonov et al., 2020). Here we bring these perspectives together to test whether human vision likewise behaves in

accordance with TWCO principles. To test this hypothesis, we used a figure-ground segregation paradigm wherein human observers reported the orientation of a rectangular figure region in a texture stimulus composed of Gabor annuli (see Figure 1c for an illustration). According to TWCO, synchrony is governed by the interplay between oscillator detuning and coupling strength (Acebrón et al., 2005). Therefore, we created stimuli in which these two core parameters were systematically manipulated. We parametrically varied contrast heterogeneity as an implementation of frequency detuning and grid coarseness as an implementation of the cortical distance that determines coupling strength (see Figure 1d for an illustration). The figure was defined by a less heterogeneous contrast distribution between the elements, compared to elements in the background. Additionally, we investigated whether this synchrony-based grouping mechanism is adaptive by using a perceptual learning paradigm in which participants improved their perceptual performance over 8 daily sessions. By formalizing the principles of TWCO in a V1 oscillator model augmented with a simple Hebbian learning mechanism, we derived quantitative predictions from the theory. Our psychophysics results align well with the synchrony exhibited by the model, supporting the idea that stimulus-dependent gamma synchrony may be behaviorally relevant.

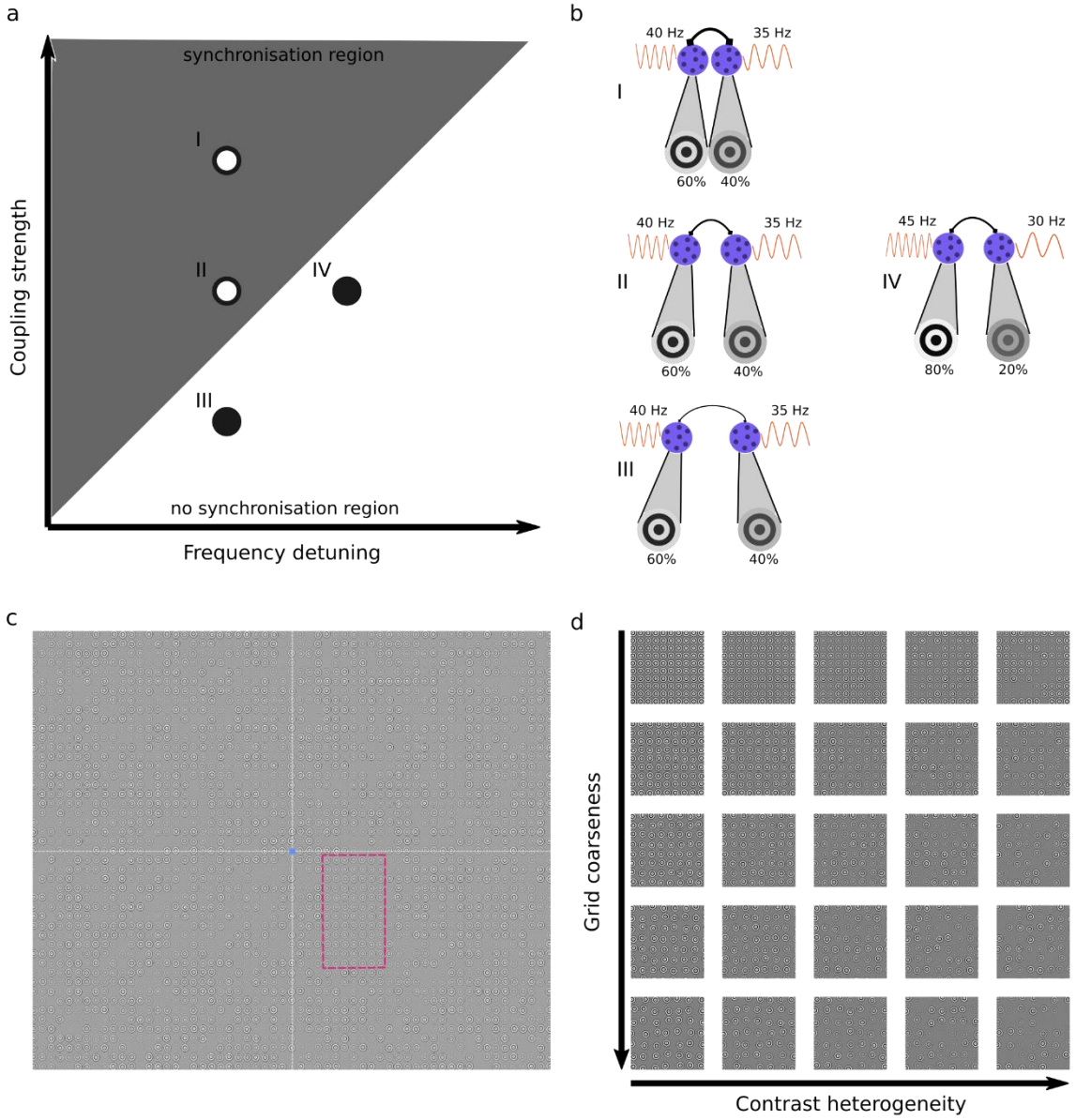


Figure 1: Schematic illustration of synchronization principles in visual cortex and stimulus design. a, Arnold tongue: triangular region shows combinations of detuning and coupling strength that allow synchrony (light grey). Open circles indicate two scenarios conducive to synchrony. The first scenario (I) combines strong coupling with moderate detuning. The second scenario (II) combines moderate coupling with moderate detuning. Closed circles indicate two scenarios not conducive to synchrony. The third scenario (III) combines weak coupling with moderate detuning. The fourth scenario (IV) combines moderate coupling with large detuning. **b**, Translation of the four scenarios to stimulus features. Detuning and coupling strength map onto contrast heterogeneity and grid coarseness through the anatomy and physiology of early visual cortex. In this simplified illustration, two texture elements (Gabor annuli) fall within receptive fields of neural assemblies (purple) in early visual cortex. Contrast determines oscillation frequency (orange), with higher contrasts leading to higher frequencies. Differences in contrast within the receptive fields of two neural assemblies thus leads to differences in their frequencies and hence to higher detuning. Note that neural assemblies typically have several Gabor annuli in their receptive fields and extract their average contrast. There is thus no one-to-one mapping between annuli and receptive fields in our model. Coupling strength (line thickness of connecting arrow) depends on cortical distance, which due to retinotopy directly relates to the distance between texture elements in the visual field. Larger distances between annuli thus stimulate more remote neural assemblies with weaker coupling. **c**, Example full texture stimulus comprised of nonoverlapping Gabor annuli on irregular grid. For all participants and sessions 1-8, the lower right quadrant contains a vertical figure (magenta outline, not shown to participants). Blue dot: fixation point. Axes separating quadrants shown for illustration only, not

visible to participants. On a given trial, the figure may be vertical or horizontal and participants indicated the figure's orientation. **d**, Figure region cut-outs illustrating experimental conditions. Grid coarseness (five steps) manipulates coupling strength for both figure and background. Contrast heterogeneity (five steps) manipulates detuning within figure. Background always at maximum heterogeneity (equivalent to rightmost column). The 25 cut-outs show all combinations of grid coarseness and contrast heterogeneity used in the experiments.

Results

Eight participants (6 female, mean age $= 23.75$, standard deviation $= 6.453$) performed a two-alternative forced choice texture discrimination task. We employed a repeated-measures design with extensive sampling. A design analysis indicated that our sample size afforded approximately 92% posterior detection probability (analogous to statistical power) for the core effects (Supplementary Table 1). Texture stimuli consisted of nonoverlapping Gabor annuli on an irregular grid (see Figure 1c). Each Gabor annulus was characterized by its own local contrast and was equiluminant with the background. Within a single visual quadrant, a rectangular figure was defined by less heterogeneity in the contrasts of local Gabor elements compared to the background, while keeping mean contrast between figure and background equal. Participants indicated the orientation (horizontal vs vertical) of the figure while fixating centrally. We manipulated two factors. The first was contrast heterogeneity within the figure, which we operationalized as the width of a uniform distribution from which annulus contrast values were drawn. This distribution was centered around a mean contrast of 50%. The background exhibited maximum contrast heterogeneity (from 0% to 100%). The second factor was the coarseness of the grid (distance between annuli). This manipulation affected figure and background equally. Both factors were manipulated in five steps resulting in 25 conditions (see Figure 1d). Within an experimental session, participants completed 30 blocks of each condition (750 trials). Participants received feedback after each trial in the form of color changes of the fixation point. Eye-tracking was used to ensure fixation, and trials where fixation was broken during either the fixation period preceding the stimulus, or during stimulus presentation, were aborted and repeated at a randomly chosen time later in the session. The experiment consisted of 9 consecutive sessions (8 training and 1 transfer session). In the transfer session, the rectangular figure was moved to the diagonally opposite quadrant.

To provide a mechanistic link between contrast heterogeneity, grid coarseness and synchrony in early visual cortex on the one hand, and quantitative predictions of discrimination accuracy on the other, we developed a phase-oscillator model of V1. The model represents a patch of visual space corresponding to the figure region in our psychophysics experiments, mapped onto V1 using a complex-logarithmic topographic transformation (Balasubramanian & Schwartz, 2002; Schwartz, 1980). To reduce computational cost, we only modeled the figure and not the background, under the assumption that the synchrony level in one image region would not be substantially altered by the synchrony level in the other image region (see Supplementary Figure 1). Based on this, synchrony in the background at maximum contrast disparity was equated to synchrony in the figure at that contrast disparity. Each model oscillator represents a neural assembly receiving local input from the visual field. The frequency of each oscillator is a quasi-linear function of the contrast falling inside its receptive field, as has been determined previously in macaques (Evers et al., 2021; Roberts, Lowet, Brunet, TerWal, Tiesinga, Fries, & DeWeerd, 2013). Receptive fields are modelled as isotropic 2D Gaussian functions with sizes that scale with eccentricity according to human cortical magnification (Freeman & Simoncelli, 2011). Furthermore, we included recurrent connections between phase oscillators reflecting the lateral anatomical connectivity among columns in V1 and other low-level visual areas (Crist et al., 2001). In line with anatomical data (Amir et al., 1993; Eckhorn, 1994; Gilbert & Wiesel, 1989; Ts'o et al., 1986), coupling strength in our model declines exponentially with physical distance along the cortical surface. Our model captures this with two parameters estimated from independent neurophysiological data (Lowet, Roberts, Peter, Gips, & De

Weerd, 2017), Peter, Gips, & De Weerd, 2017): maximum coupling strength γ and coupling decay factor λ. The model was exposed to the same figure region texture stimuli as human participants, with manipulations of contrast heterogeneity and grid coarseness. We quantified the model's degree of zero-lag synchrony as the magnitude of the Kuramoto order parameter (synchronization index).

In our V1 model, learning is implemented to occur offline between simulated sessions, following a Hebbian-type learning rule that adapts coupling strengths based on pairwise phase-locking values (PLVs) accumulated over trials within a session. The contribution of each trial to learning is weighted by the probability of a correct response, determined by a psychometric function relating model synchrony to performance. This learning mechanism implies that connections between oscillators that exhibited coherence on correct trials are strengthened, bounded by the maximum coupling strength. Incorporating an upper bound on connections was motivated by findings that synaptic strength is limited by intrinsic properties of vesicular docking (Malagon et al., 2020) and that late long-term potentiation approaches a maximum after several repeated experiences (Kandel et al., 2000). Free parameters of the learning mechanism were estimated using data from the first two experimental sessions. To maximally disentangle data used for adjusting model parameters and data used for testing model predictions, we employed a leave-one-out cross-validation procedure. Model parameters were repeatedly estimated from the first two sessions in seven of our eight participants and the resulting model was used to predict performance in the remaining six sessions of the left-out participant. Our model rests on the assumption that learning-induced structural changes in early visual cortex are specific to the retinotopic locations of the trained stimuli. We evaluated whether this assumption holds for our human participants using the transfer session following the main training period. In the transfer session, participants performed the texture discrimination task with the figure region moved to a visual quadrant that had not been previously exposed to the figure. If learning is indeed local, participants' performance in the transfer session should resemble that of early training sessions, indicating a reset in performance for the new retinal location. On the other hand, if learning generalizes across retinal locations, performance in the transfer session should maintain the improvements seen in later training sessions. By comparing transfer session performance to both early and late training sessions, we can evaluate the validity of our model's assumption.

Synchrony Principles Govern Static Figure-Ground Perception

We first asked the question whether the factors that determine synchrony among coupled oscillators, frequency detuning and coupling strength, are predictive of the human ability to segregate a rectangular figure from its background in texture stimuli. In early visual cortex, oscillation frequency directly maps onto the contrast of texture elements (Hadjipapas et al., 2015; Lowet et al., 2015; Roberts, Lowet, Brunet, TerWal, Tiesinga, Fries, & DeWeerd, 2013) and coupling strength directly maps onto their physical proximity (Gilbert & Wiesel, 1983; Lowet et al., 2015; Lowet, Roberts, Peter, Gips, & De Weerd, 2017; Stettler et al., 2002; Ts'o et al., 1986). If texture segregation indeed depends on the synchrony principles identified by the theory of weakly coupled oscillators (TWCO), we expect discrimination accuracy to reveal a "behavioral" Arnold tongue in the space defined by contrast heterogeneity and grid coarseness.

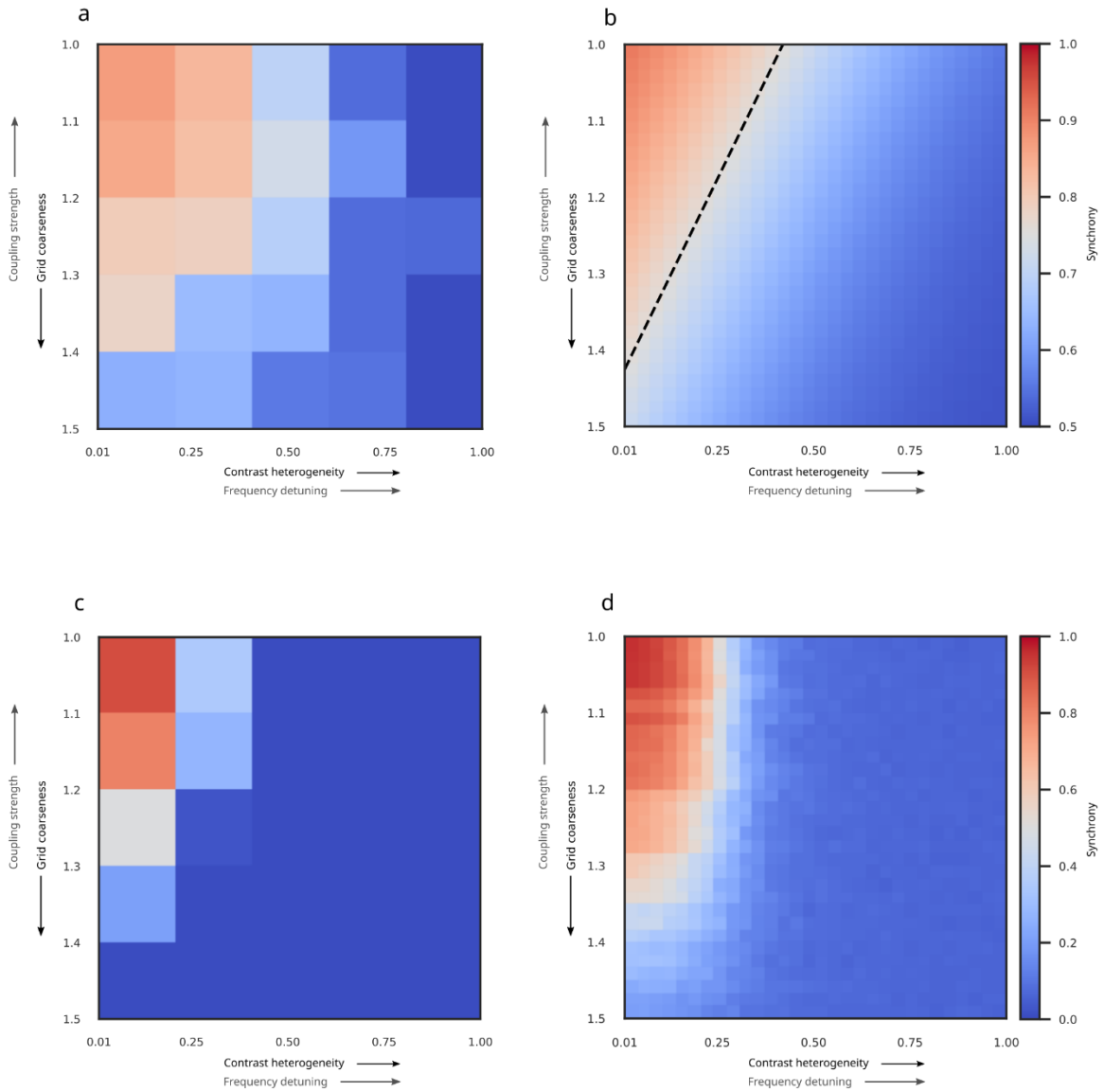


Figure 2: Behavioral and simulated Arnold tongues. a, Average discrimination accuracy for each of the 25 experimental conditions revealed a behavioral Arnold tongue in the space defined by contrast heterogeneity and grid coarseness. Contrast heterogeneity translates into the variance of frequencies (detuning) whereas grid coarseness translates into cortical distance (coupling strength). **b**, Fitted behavioral Arnold tongue after fitting a two-dimensional psychometric curve to the results in (a). The dashed line indicates the combination of contrast heterogeneity and grid coarseness corresponding to 75% accuracy. **c**, Zero-lag synchrony among model oscillators showing an Arnold tongue in the same parameter space as (a). Simulation conditions matched the 25 experimental conditions. **d**, High-resolution visualization of zero-lag synchrony, using 900 conditions (30 levels each of contrast heterogeneity and grid coarseness) to provide a more detailed representation of the Arnold tongue.

To test these predictions, we analyzed the main effects of contrast heterogeneity and grid coarseness, as well as their interaction, on discrimination accuracy using Bayesian hierarchical logistic regression. This allowed us to analyze individual trial data rather than aggregated accuracy, while simultaneously accounting for within-subject variability by estimating participant-specific intercepts and slopes for each predictor. Both contrast heterogeneity and grid coarseness were z-normalized prior to fitting the statistical model. Note that while the principles of TWCO primarily predict main effects of contrast heterogeneity and grid coarseness, we additionally included their

interaction to capture complex relationships specific to V1 that are not immediately apparent from the general theory. Specifically, coupling strength decays exponentially with cortical distance, which itself depends on cortical magnification. This should lead to a highly nonlinear relationship between grid coarseness and coupling strength that is likely to manifest as an interaction. In line with our expectations, the test provided strong evidence that both increased contrast heterogeneity (β =−0.60, 95% HDI [−0.89, −0.30], Pr[β<0] = 0.999, OR = 0.56, 95% HDI for OR [0.41, 0.74]) and grid coarseness (β = −0.27, 95% HDI [−0,40 −0.13], Pr[β<0] = 0.999, OR = 0.77, 95% HDI for OR [0.67, 0.88]) reduced discrimination accuracy (Figure 2a,b). These results provide credible evidence that a one-standard-deviation increase in contrast heterogeneity reduces the odds of a correct response by approximately 44%, while a similar increase in grid coarseness reduces the odds by 23%. Furthermore, Figure 2a,b shows a behavioral Arnold tongue as a triangular region of high accuracy (≥75% correct). There was likewise strong evidence for an interaction between contrast heterogeneity and grid coarseness (β = 0.24, 95% HDI [0.12, 0.36], Pr[β>0] = 0.998, OR = 1.27, 95% HDI for OR [1.12, 1.44]). This indicates that the specific characteristics of early visual cortex contribute beyond the general principles of TWCO.

Our model of V1 captures both the general principles of TWCO as well as idiosyncratic characteristics of early visual cortex in a single mechanism, and we expected this model to predict the human ability to segregate a rectangular figure from its background in texture stimuli. Indeed, the synchrony exhibited by our model (Figure 2c,d), when exposed to the same stimuli as our participants, resembled behavioral discrimination accuracy (Figure 2a,b). A Bayesian hierarchical logistic regression with model synchrony as sole predictor revealed strong evidence that it is associated with improved accuracy (β = 0.76, 95% HDI [0.33, 1.21], Pr[β>0] = 0.998, OR = 2.19, 95% HDI for OR [1.38, 3.33]). This represents credible evidence that a one-standard-deviation increase in synchrony more than doubles the odds of a correct response. Hence, our proposed mechanism is capable of reproducing the key patterns in the behavioral data.

A natural question is whether synchrony constitutes the unique mechanistic link from stimulus features to perception within our model. To address this question, additional analyses used the average model firing rates within the figure as a predictor for segregation, as well as the difference between average model firing rates inside and outside the figure. The latter rate difference between figure and ground can serve as a phenomenological proxy for putative rate-based segregation mechanisms. Note that we treat the instantaneous frequency of each oscillator as a proxy for the instantaneous population firing rate of the corresponding neural assembly.

With respect to the average figure firing rates, we found some evidence indicating that they were associated with segregation accuracy (β = 0.07, 95% HDI [-0.025, 0.16], Pr[β>0] = 0.941, OR = 1.07, 95% HDI for OR [0.98, 1.18]). However, because the 95% highest density interval included zero, we evaluated whether the effect fell within a Region Of Practical Equivalence (ROPE) of ±2% accuracy and found only weak evidence for this (Pr[|Δacc|<0.02] = 0.680). Hence, the effect is likely present, but small. With respect to rate differences, we found credible evidence that they could be associated with accuracy (β = -0.55, 95% HDI [-0.78, -0.32], Pr[β<0] = 0.999, OR = 0.58, 95% HDI for OR [0.46, 0.73]). However, the effect of rate difference was smaller than that of synchrony. Furthermore,firing in the figure was reduced compared to background firing.

We next compared synchrony, average figure firing rate and rate differences derived from the same V1 simulations in terms of their out-of-sample predictive accuracy using Pareto-smoothed importance sampling leave-one-out cross-validation. Synchrony was favored over rate difference (ΔELPD ≈ 19, dSE ≈ 14) and average figure firing (ΔELPD ≈ 127, dSE ≈ 17). The stacking weights further support this with synchrony receiving a weight of 0.90, rate difference a weight of 0.10, and the average figure firing rates a weight of effectively zero. These results indicate that, for our stimuli and our V1 model, a synchrony-based readout provides the most faithful mapping from stimulus to

perception among simple alternatives. However, this comparison does not rule out that more sophisticated rate-based models could provide viable mechanistic accounts of figure-ground segregation. Nevertheless, our data indicate that synchrony-based mechanisms are eminently viable.

A key strength of our model is that it does not depend on fine-tuning parameters to our behavioral data. To demonstrate this, we conducted a parameter space exploration of key choices of model parameter values (maximum coupling strength and coupling decay factor) and found that our choices, which were obtained from independent observations in macaques (Lowet, Roberts, Peter, Gips, & De Weerd, 2017) were already close to optimal. We used Pearson correlations (Figure 3a) and weighted Jaccard similarity (Figure 3b) to assess the similarity between the behavioral Arnold tongue and the Arnold tongue predicted by our V1 model for various combinations of maximum coupling strength and coupling decay factor. We included both correlations and Jaccard similarity because the former is more widely known while the latter is more conservative. To compute weighted Jaccard similarity between two sets of real numbers, they need to fall within the same range. Accordingly, we applied min-max normalization to ensure that discrimination accuracy fell within a zero-to-one range matching the range of the synchronization index. This procedure yielded the similarity comparisons color-coded in Figure 3. The point marked with the black dot reflects the parameter value combination that was based on independent macaque data (24.63 and 0.22, respectively) and that was exclusively used for our model predictions. When using Pearson correlations as a similarity measure, this parameter value combination fell just within the region of optimal parameter values for our behavioral results (Figure 3a). When using the more conservative weighted Jaccard similarity index (Figure 3b), our chosen parameter value combination appeared slightly outside of the optimal region. Thus, the two model parameters estimated from neurophysiological recordings in monkeys were close to optimal for predicting human perceptual behavior, but not fully optimal. This may be due to horizontal connections in human visual cortex extending further than those in the macaque (Amir et al., 1993; Burkhalter & Bernardo, 1989; Lund, Yoshioka, & Levitt, 1993; Voges et al., 2010; Yoshioka et al., 1996), suggesting a slightly smaller coupling decay factor in humans. Applying a smaller coupling decay would move model parameters into the optimal regime, thereby extending the predicted Arnold tongue (Figure 2c) diagonally in the direction of the behavioral Arnold tongue (Figure 2a). The parameters maximum coupling and decay factor might reflect biological constraints on the strength of lateral connections (Kandel et al., 2000; Malagon et al., 2020; Rioult-Pedotti et al., 1998), which to some extent may differ between monkeys and humans.

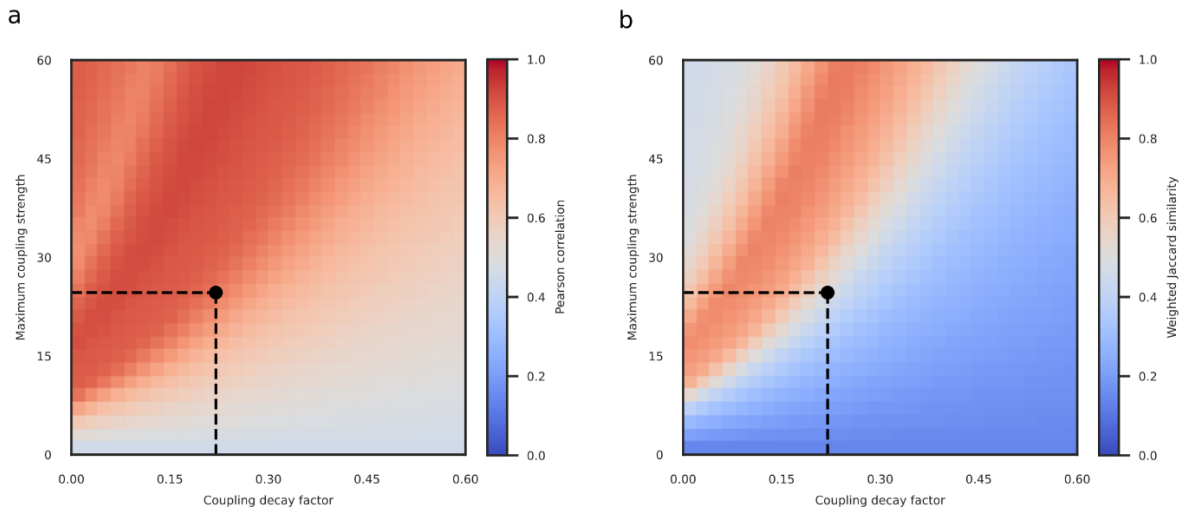


Figure 3: Comparison of behavioral and simulated Arnold tongues across coupling parameter space. a, Pearson correlation between the behavioral Arnold tongue and simulated Arnold tongues obtained from models with coupling weights determined by different combinations of maximum coupling strength and coupling decay factor. The point labelled by the black circle shows the combination of parameters that were obtained from independent (macaque) data. **b**, Weighted Jaccard similarity between the behavioral Arnold tongue and simulated Arnold tongues. This metric is displayed across the same parameter space as in (a).

Plasticity-Induced Changes in Synchrony Quantitatively Predict Perceptual Learning

Our results show that a neural grouping mechanism based on synchrony principles can account for behavioral performance, suggesting it is a viable candidate for explaining texture segregation. We next asked whether training-induced changes of lateral connections among neural assemblies in early visual cortex affect assemblies' readiness to synchronize and whether this is accompanied by performance improvements. We reasoned that neural synchrony must remain adaptable to the statistics of visual experiences to function effectively as a grouping mechanism. Consequently, we hypothesized that if synchrony among neural assemblies is related to figure-ground segregation and enhanced through perceptual learning, the ability to segregate figure from ground should increase with training.

To test this, both the model and human participants were exposed to eight daily sessions of extensive training using identical stimuli and experimental conditions. We hypothesized that both grid coarseness and contrast heterogeneity exhibit main effects on discrimination accuracy. However, we also expected coupling strength to increase with learning and that this would allow synchrony to occur for increasingly coarser grids. We thus hypothesized an interaction effect between session and grid coarseness on discrimination accuracy. Furthermore, we hypothesized an additional interaction between session and contrast heterogeneity where the effect of contrast heterogeneity would increase over sessions. Model simulations for the first session never revealed a synchronized state for contrast heterogeneity values beyond 0.25, even for the densest grids (see Figure 2c,d). This, together with an upper bound on coupling strength, suggested that synchrony cannot be achieved far beyond this cutoff point, even after extensive training. Indeed, model

simulations of training confirmed this, showing that synchrony approached this cutoff point for increasingly coarser grids over sessions (Figure 4c). These model results indicate that the effect of contrast heterogeneity would increase over sessions with high performance for values below the cutoff point and low performance above the cutoff point. Finally, extensive training may globally increase participants' performance, implying a main effect of session.

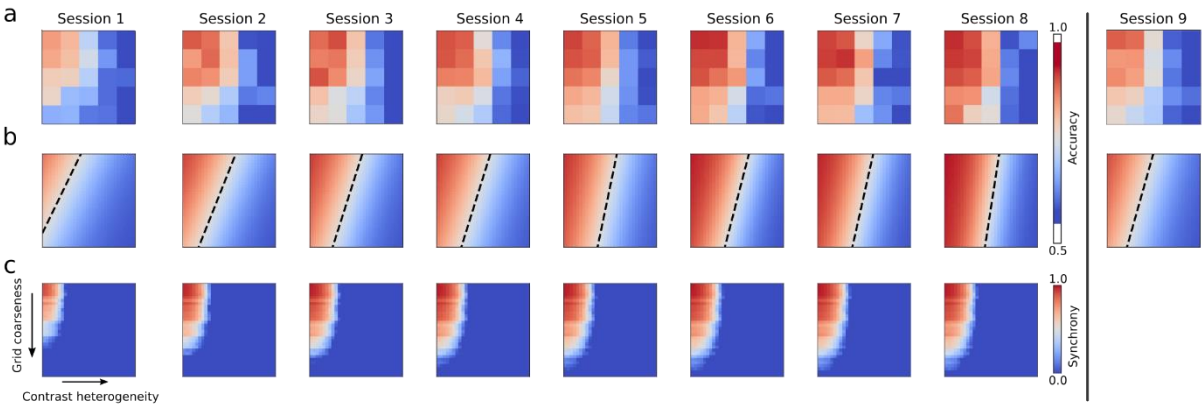


Figure 4: Learning effects on Arnold tongues. a, Group average behavioral Arnold tongues for the 25 experimental conditions for each session. The vertical black line separates transfer session 9 from training sessions 1 to 8. **b**, Two-dimensional psychometric curves fitted to session-specific group average behavioral Arnold tongues. The dashed line again indicates the combination of contrast heterogeneity and grid coarseness at which participants achieve 75% accuracy. **c**, Simulated Arnold tongues for each of the eight training sessions including session-by-session learning in the model. We did not include a simulation of the ninth session because the location-specificity of the model learning rule would render it identical to the first session. Note that for visualization purposes we simulated the model for 30 levels of contrast heterogeneity and 30 levels of grid coarseness, in both cases including the 5 levels investigated experimentally.

To evaluate these predictions, we performed Bayesian hierarchical logistic regression including the main effects of contrast heterogeneity, grid coarseness, and session, as well as interactions between session and contrast heterogeneity, between session and grid coarseness, and between contrast heterogeneity and grid coarseness. As before, we included the interaction between contrast heterogeneity and grid coarseness to account for potential nonlinear effects specific to V1. The analysis revealed strong evidence that participants' ability to segregate figure from ground increased over sessions ($\beta = 0.095$, 95% HDI [0.066, 0.126], $Pr[\beta>0] = 1.00$, OR = 1.10, 95% HDI for OR = [1.07, 1.13]). Furthermore, we found strong evidence for an interaction between contrast heterogeneity and session ($\beta = -0.081$, 95% HDI [−0.090, −0.070], $Pr[\beta<0] = 1.00$, OR = 0.92, 95% HDI for OR = [0.91, 0.93). We also found evidence for a small interaction between grid coarseness and session ($\beta = -0.008$, 95% HDI [−0.018, 0.001], $Pr[\beta<0] = 0.954$, OR = 0.99, 95% HDI for OR = [0.98, 1.00]). However, the 95% highest density interval included zero. We subsequently confirmed that the change was within a region of practical equivalence (ROPE) of ±2% accuracy $Pr[|\Delta acc|<0.02] = 0.999$. While the interaction between session and grid coarseness is thus negligible, there was strong evidence for a main effect of grid coarseness on discrimination accuracy ($\beta = -0.316$, 95% HDI [−0.377, −0.262], $Pr[\beta<0] = 1.00$, OR = 0.73, 95% HDI for OR = [0.69, 0.77]) with increasing accuracy as grid coarseness decreased. As can be appreciated from Figure 4, there indeed seemed to be a cutoff value for contrast heterogeneity beyond which the figure could not be discriminated from the background. This cutoff may also explain why the interaction between session and grid coarseness was negligible. Below the cutoff point, the top and middle rows of Figure 4 suggest that participants could discriminate the figure for increasingly coarser grids. Beyond the cutoff, however, grid coarseness seemed to have had no discernible effect regardless of how much training participants received. The characteristic triangular shape of the Arnold tongue thus gradually morphed into a rectangular shape.

Next, we examined simple effects of contrast heterogeneity on discrimination accuracy for each session separately (see Table 1). As expected from the presence of the cutoff, the effect of contrast heterogeneity increased over sessions, reflected in decreasing log-odds (β) and corresponding odds ratios (ORs) over sessions as shown in Table 1.

Table 1: Effects of contrast heterogeneity on discrimination accuracy across sessions. HDI = Highest Density Interval. Log-odds and Odds Ratios represent the effect of a one standard deviation increase in contrast heterogeneity on the odds of correct discrimination. The probabilities of negative log-odds are for the simple effect of contrast heterogeneity in each session.

Session	Log-odds (β)	Odds Ratio (OR)	95% HDI for OR	Pr[β<0]
1	-0.66	0.52	[0.38, 0.66]	1.00
2	-0.74	0.48	[0.35, 0.61]	1.00
3	-0.82	0.44	[0.32, 0.56]	1.00
4	-0.90	0.41	[0.30, 0.52]	1.00
5	-0.98	0.38	[0.28, 0.48]	1.00
6	-1.06	0.34	[0.26, 0.44]	1.00
7	-1.14	0.32	[0.24, 0.41]	1.00
8	-1.22	0.30	[0.22, 0.38]	1.00

Finally, we evaluated whether the observed effects reflected localized learning in early visual cortex, as assumed by our model, implying that the training effect would be specific to the trained location. Performance in the transfer session should thus resemble that observed at training locations during early rather than late sessions. To test this, we estimated a hierarchical Bayesian logistic regression model with predictors for contrast heterogeneity, grid coarseness, session, their interactions, and an indicator for the transfer session. Subject-level random intercepts and slopes were included. From the fitted model, we generated posterior predictions of the population-level mean accuracy for each session. We then compared transfer (Session 9) with an early reference session in two complementary ways. First, we estimated the posterior probability that transfer session accuracy was lower than in the reference session. Second, we estimated the posterior probability that the difference between accuracy in the transfer and reference session lay within a region of practical equivalence (ROPE, ±2% accuracy). We used the second session, the earliest session after task familiarization, as reference. Our analysis revealed that performance in the transfer session was practically equivalent to session 2 (93% posterior probability of equivalence). Based on this, we expected that the local learning mechanism implemented in our model can provide quantitative predictions of performance changes over the course of the eight training sessions.

We evaluated the quantitative agreement between model synchrony and empirical discrimination performance. This analysis focused exclusively on synchrony. As we show in the supplementary materials (Suppl. Figure 2), rate-based readouts of our V1 model are not at all affected by coupling strength. As such, they are insensitive to changes in coupling and are thus not viable as alternative mechanisms to explain performance changes due to learning. To evaluate the quantitative agreement between model synchrony and empirical discrimination performance, we measured the similarity between simulated and behavioral Arnold tongues using Pearson correlations and weighted Jaccard similarity. Because we employed a leave-one-out cross-validation procedure, we obtained eight simulated Arnold tongues in sessions 2-8 after optimizing learning parameters on data from seven participants. Simulated Arnold tongues in each fold were always compared to behavioral Arnold tongues of the left-out participant. The first session did not involve learning and model simulations were identical to those reported above. Note that data from the second session was used to adjust model parameters and hence only sessions 3-8 could be used for evaluating model predictions. This cross-validation approach enabled us to assess the model's ability to predict performance in unseen data, rather than merely fitting observed results post-hoc. Figure 5a and 5b show correlations and Jaccard similarity between simulated and behavioral Arnold tongues, respectively. Grey regions indicate a noise ceiling that was obtained by computing the fit between average behavioral Arnold tongues in a fold and the behavioral Arnold tongue of the left-out participant. The grey region marks the 25th to the 75th percentile of fit values obtained using this procedure. The figure demonstrates consistent quantitative agreement between simulated and behavioral Arnold tongues across sessions.

To examine this further, we tested whether the size of the simulated Arnold tongue across sessions was predictive of the size of the behavioral Arnold tongues. We quantified the size of each Arnold tongue in terms of the volume under its surface computed using Simpson's numerical integration. Arnold tongues grew across sessions with comparable growth curves for simulated and behavioral Arnold tongues (see Figure 5c). The precise relationship between simulated and behavioral Arnold tongue sizes is depicted in Figure 5d. Subsequently, we performed a Bayesian hierarchical linear regression to investigate this in sessions 3-8. We ignored the first two sessions since these were used for estimating model learning parameters. As expected, the size of the simulated Arnold tongue predicted the size of the behavioral Arnold tongue (β = 0.54, 95% HDI [0.106, 0.935], Pr[β>0] = 0.992). The model's capability to accurately reflect learning effects observed in human participants is consistent with the notion that enhanced synchrony among neural assemblies in early visual cortex resulting from perceptual learning enhances human's ability to segregate figure from ground. This further strengthens the view that synchrony principles provide a viable neural grouping mechanism for texture segregation.

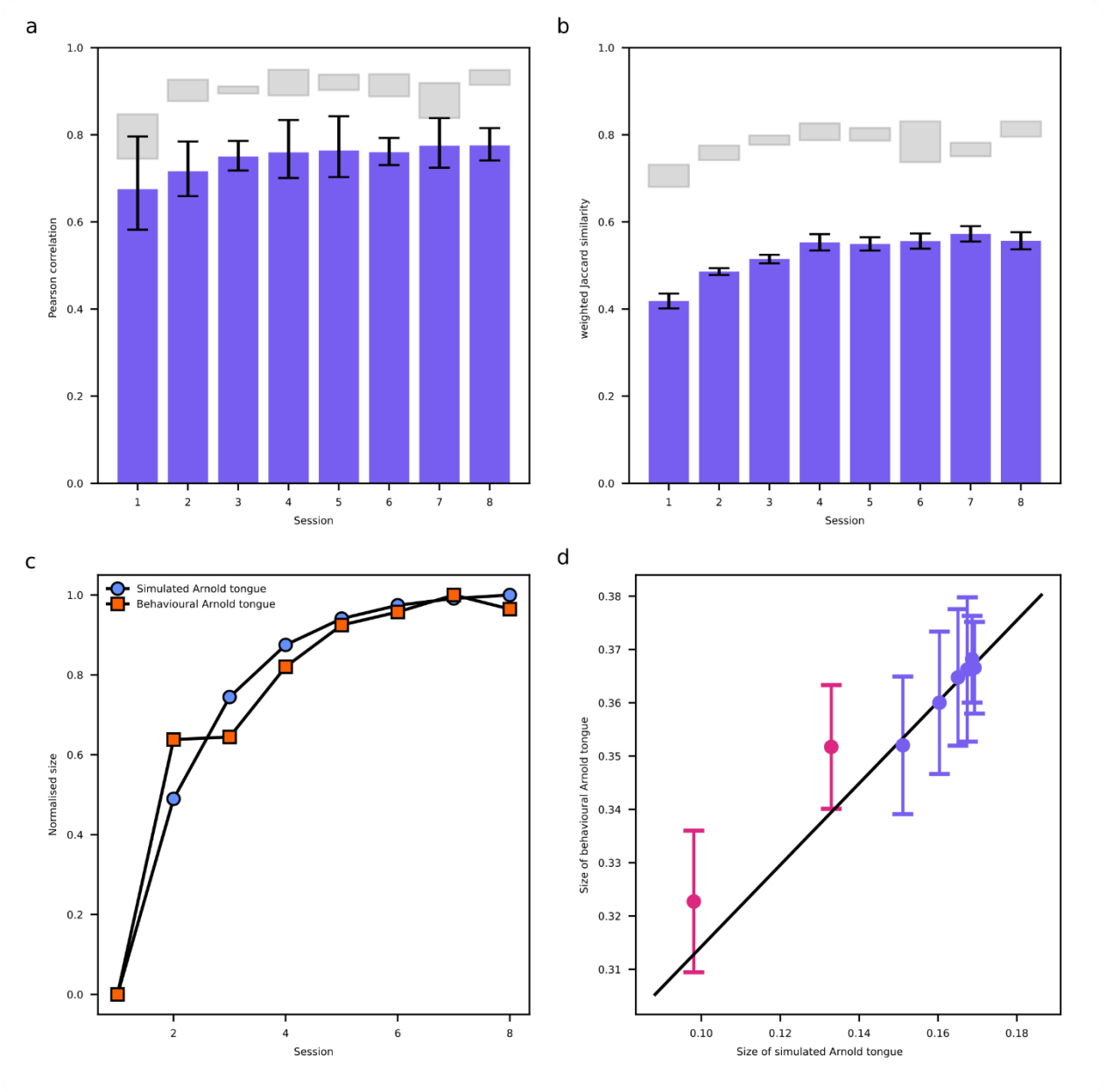


Figure 5: Model predictions of learning effects. a, Pearson correlations between simulated and behavioral Arnold tongues for each training session. Error bars indicate 95% confidence intervals. Grey regions indicate a noise ceiling that was obtained by computing the fit between average behavioral Arnold tongues in a fold and the behavioral Arnold tongue of the left-out participant. The grey regions reflect the 25th to the 75th percentile of fit values obtained using this procedure. **b**, Weighted Jaccard similarity values between simulated and behavioral Arnold tongues for each training session. Error bars and grey regions as in (a). **c**, Sizes of simulated (blue circles) and behavioral (orange squares) Arnold tongues across sessions. Arnold tongue sizes were averaged across participants and subsequently min-max normalized. This normalization highlights the growth patterns while accounting for the different value ranges of simulated and behavioral Arnold tongues. **d**, Sizes of behavioral Arnold tongues as a function of sizes of simulated Arnold tongues. The best fitting regression (black line) was obtained from a mixed effects model fitted to data from sessions 3-8 (blue circles). Red circles reflect data from the first two sessions that was not included in the mixed effect model. The black line was extended to include these points. Error bars indicate 95% confidence intervals.

Discussion

The role of synchrony in the gamma frequency band for visual perception remains a matter of debate (Duecker et al., 2021; Fernandez-Ruiz et al., 2023; Ray & Maunsell, 2015; Roelfsema, 2023). A putative role for gamma synchrony in processing the features of a stimulus both within and across visual areas (Fries, 2009; Singer, 1999; Uhlhaas et al., 2008; Womelsdorf et al., 2007) has been called into question based on the stimulus-dependence of gamma synchrony (Ray & Maunsell, 2010, 2015; Roelfsema, 2023). Alternatively, it has also been suggested that feature-dependent gamma frequencies and distance-dependent synchrony are key ingredients in a neural grouping mechanism underlying figure-ground segregation (Lowet et al., 2015; Lowet, Roberts, Peter, Gips, & De Weerd, 2017). It is well-established that the frequency of gamma oscillations in visual cortex depends on local stimulus features (Baldi & Meir, 1990; Buia & Tiesinga, 2006; Hall et al., 2005; Henrie & Shapley, 2005; Roberts, Lowet, Brunet, TerWal, Tiesinga, Fries, & DeWeerd, 2013; Shapira et al., 2017) and that lateral connectivity between neural groups within early visual cortex depends on cortical distance (Amir et al., 1993; Boucsein et al., 2011; Eckhorn, 1994; Gilbert & Wiesel, 1989; Stettler et al., 2002; Ts'o et al., 1986). It is likewise a well-known property of coupled oscillators that they synchronize when their coupling is sufficiently strong to overcome differences in their frequency, but not otherwise (Acebrón et al., 2005; Ermentrout et al., 2019; Kuramoto, 1984; Neu, 1979; Strogatz, 2000). Synchrony may thus drive the perceptual grouping of elements if they are sufficiently similar to each other within one image region, and thereby segregate it from other image regions based on their different levels of synchrony.

We tested this hypothesis in a psychophysics experiment wherein human observers discriminated the orientation of a texture-defined, rectangular figure region (vertical vs horizontal). The stimulus consisted of small Gabor annuli arranged on an irregular grid. Each Gabor annulus was characterized by its own contrast and the figure region was defined by less heterogenous contrasts among the Gabor annuli compared to the background. We manipulated contrast heterogeneity and grid coarseness (distance between annuli) as a proxy of frequency detuning and coupling strength, respectively. Both contrast heterogeneity and grid coarseness affected discrimination accuracy. Specifically, we found that accuracies beyond 75% were limited to a triangular region in the space spanned by these two factors, forming a behavioral Arnold tongue. In line with our expectations, increased contrast heterogeneity in the figure permitted figure-ground segregation if accompanied by a reduction in grid coarseness. These results quantitatively aligned well with synchrony exhibited in a coupled oscillator V1 model exposed to the same texture stimuli.

The capacity of our model to predict human psychophysical performance is notable given that the key parameters of maximum coupling strength and coupling decay factor were obtained from neurophysiological data recorded from macaques (Lowet, Roberts, Peter, Gips, & De Weerd, 2017). This cross-species validation underscores the robustness of our model and suggests that the neural mechanisms underlying gamma oscillations and figure-ground segregation are largely conserved across primate species (Buzsáki et al., 2013). It is noteworthy that the parameter combination obtained from macaque data bordered the region of optimal combinations exhibiting the highest match to psychophysics results that our model could, in principle, achieve (see Figure 3). The slight deviation from the optimal regime likely stems from the fact that parameters were estimated from data obtained in macaques and subsequently used to predict human behavior. It is likely that horizontal connections in the human extend further than those in the macaque (Amir et al., 1993; Burkhalter & Bernardo, 1989; Lund, Yoshioka, & Levitt, 1993; Lund, Yoshioka, Lund, et al., 1993; Voges et al., 2010) and may thus be associated with a slightly smaller coupling decay factor. Another possibility is that the parameter value we derived from Lowet et al. (2017), a study chosen because their paradigm targets the same TWCO components that guided our stimulus design, is an overestimate. As with any study, their data comes with uncertainty such that our estimates might not perfectly reflect actual decay rates. While we currently do not have alternative data to estimate

the exact human decay factor and hence cannot establish how much model fit would be affected, any small to modest reduction would certainly further improve model fit.

To further investigate whether synchrony among neuronal populations exhibiting contrast dependent frequencies provides a potential perceptual grouping mechanism, we tested whether training-induced changes of lateral coupling in a network of phase oscillators improved the readiness of these oscillators to synchronize and whether this model provided accurate predictions of performance on the figure-ground segregation task. We reasoned that for neural synchrony to function effectively as a grouping mechanism, it should be modifiable by experience in a manner that matches training-induced improvements in texture segregation. We observed that discrimination performance improved as a function of training session, in line with participants' growing experience with the stimuli. Importantly, we found that training-induced increases in accuracy were well accounted for by model predictions of synchrony strength inside the figure. Our results are consistent with the notion that synchrony mechanisms in low-level visual areas contribute in a behaviorally relevant manner to texture segregation and that training-induced changes of local synchrony are reflected by concurrent changes in perception. Synchrony and discrimination accuracy revealed highly congruent Arnold tongues. A close quantitative resemblance of these Arnold tongues was maintained across sessions as both tongues grew and changed form in a highly consistent manner. This supports the idea that learning-induced changes in figure-ground segregation may be mediated by plasticity-induced changes in synchrony. Oscillations have been shown to facilitate learning through spike-timing dependent plasticity (Masquelier et al., 2009), rendering an oscillation-based Hebbian learning mechanism biologically plausible.

It is important to note that the learning mechanism integrated into our model assumes that learning is local. We validated this assumption in the human participants by testing whether moving the figure region from its trained location to a new location would lead to transfer of performance to the new location, or rather a decrease in performance in the new location. Our results supported the latter. This is in line with other studies that demonstrated that after location-specific training, low-level visual areas contribute to the location and stimulus specificity of expert visual performance (Karni & Bertini, 1997). Based on previous findings that location-specific training induces localized plasticity in low-level visual areas (Brosch et al., 2015; Raiguel et al., 2006; Schoups et al., 2001; Yang & Maunsell, 2004), we further assumed that learning in our paradigm primarily affects lateral connectivity within V1 and hence manipulates coupling strength between neural assemblies. An alternative hypothesis could be that learning, by targeting feedforward or feedback connectivity, alters the contrast sensitivity of neural assemblies. If this were to reduce the slope of the contrast-frequency relationship, it could theoretically offer a pathway to achieve synchrony across more heterogeneous contrasts by minimizing detuning rather than increasing coupling strength. However, empirical evidence suggests that training on perceptual tasks tends to steepen, rather than flatten, the contrast-frequency relationship (Chen et al., 2013; Hua et al., 2010; Sanayei et al., 2018). Given its lack of empirical support, we therefore did not incorporate this alternative mechanism into our model.

The predominant cue for figure-ground segregation in our stimuli lay in the global variations of population statistics in the contrast distribution, rather than local differences at the boundary between the figure and the ground (de Weerd et al., 1994; Poort et al., 2016; Roelfsema et al., 2002). This design was specifically chosen to preclude simple segregation based on mean firing rates. Nevertheless, our results indicate that a region comparison mechanism could still exploit firing rates to segregate figure from ground. While the firing rates of individual oscillators in our model are modulated by ongoing interactions, average firing rates in the figure region are insensitive to these interactions and hence purely driven by feedforward contrast extraction. By contrast, synchrony in our model arises from these interactions as they convert variance in local firing rates into coherence signals. Our results show that this can provide sufficient information for a subsequent read-out mechanism to distinguish figure from background. It might also provide additional information that

downstream regions might exploit in addition to information carried by average firing rates. It might, for instance, provide a scaffold that can then be refined and read out by top-down mechanisms (Ahissar & Hochstein, 1997, 2004; Hochstein & Ahissar, 2002; Liu & Weinshall, 2000; Rubin et al., 1997). Such a scaffold might be compatible with widely accepted recurrent models in which boundary detection is followed by region-filling feedback (Grossberg & Mingolla, 1985, 1987; Keil et al., 2005; Layton et al., 2014; Motoyoshi, 1999; Neumann et al., 2001; Pessoa & de Weerd, 2003; Roelfsema et al., 2002), a notion substantiated by neurophysiological and psychophysical evidence (Poort et al., 2016; Roelfsema et al., 2002; Self et al., 2012) as well as by lesion and optogenetics experiments (Kirchberger et al., 2021; Lamme et al., 1998; Supèr & Lamme, 2007). We must note that we used the instantaneous frequencies of our model oscillators as a proxy for population firing rates. This is an oversimplification given that population firing rates are much lower than gamma (Zachariou et al., 2021). However, over the contrast range relevant to our stimuli gamma frequency and population firing co-vary approximately linearly (Zachariou et al., 2021). Frequency thus served as a rate-like activation measure rather than a literal firing rate. It is furthermore important to acknowledge that our model does not account for attentional effects, although the significance of attention in figure-ground segregation and in learning is well-established (Huang et al., 2020) and it is likely that pure exposure to the stimuli in our experiment would have revealed very limited effects (Seitz & Dinse, 2007). Thus, while the current model indicates what early visual circuits could achieve in isolation, integrating the synchrony scaffold with rate-based mechanisms and attentional gain control remains a goal for future work.

A consideration to keep in mind in interpreting the effects of training on texture segregation is that participants at the outset of the experiment were unfamiliar with various aspects of the task unrelated to the perceptual challenge itself. They had to learn to maintain fixation, to establish stimulus-response mappings and associated decision processes, in addition to solving the perceptual challenge. As such, results in the first session may represent cognitive processes related to these non-perceptual factors. Future versions of our experiment might consider including a baseline training session during which participants get acquainted with the experimental setup and task using stimuli that define figure and background with features that are independent of those manipulated in the main experiment. Moreover, participants were not informed of which visual quadrant the figure would appear in the transfer session. This raises the concern that our results partly reflect visual search effects (Eckstein, 2011; Neisser, 1964) rather than a return to a naïve state of the figure-ground segregation skill. Arguably, however, this only affected a few trials and is thus insufficient to account for the loss of skill we observed. Furthermore, our model was designed to test the emergence of synchrony within the figure region itself, and as such, it did not include the background texture. While this approach allowed us to isolate the core mechanism of interest, it means our model provides an account of local grouping rather than a full simulation of figure-ground segregation. Finally, although a strength of this work is the prediction of human psychophysical performance based on a model whose parameters were set by independent neurophysiological data, a weakness is the absence of neurophysiological data for our specific experimental paradigm. Such data would allow for a full mediation analysis from stimulus features via synchrony to behavior and could strengthen our interpretations. At the same time, a combined psychophysical and neurophysiological experiment in an animal model replicating the experimental conditions used here would benefit from strong predictions provided by the present study as well as prior neurophysiological data (Lowet, Roberts, Peter, Gips, & De Weerd, 2017). Despite these considerations and limitations, our results support the notion that gamma synchrony can serve a mechanistic role in figure-ground segregation.

The synchrony-based grouping mechanism studied here provides a theoretical framework for previous experimental results. A wide range of texture manipulations have been shown to drive segregation, including contrast (Hadjipapas et al., 2015), spatial frequency (Bredfeldt & Ringach, 2002; Henriksson et al., 2008), color (Shapley & Hawken, 2011), orientation (Lamme, 1995) and movement direction (Lamme, 1995). It is well documented that the difference between figure and

background in one or a combination of these features (Landy & Bergen, 1991; Motoyoshi & Nishida, 2001; Nothdurft, 1985a, 1991b, 1991a) in population statistics (de Weerd et al., 1992; Nothdurft, 1985b) and in the physical proximity among texture elements within a figure (de Weerd et al., 1992; Nothdurft, 1985b) are the main parameters that determine the accuracy of figure-ground segregation. Much of this work consists of separate studies focusing on the contributions of single or restricted subsets of features to segregation. Viewed through the lens of TWCO, however, these features have their effect through the same mechanism. Most element features directly influence frequency detuning (Dubey & Ray, 2020; Hadjipapas et al., 2015; Henrie & Shapley, 2005; Roberts, Lowet, Brunet, TerWal, Tiesinga, Fries, & DeWeerd, 2013; Shapira et al., 2017), while proximity determines coupling strength via lateral connectivity in early visual cortex (Boucsein et al., 2011; Gilbert & Wiesel, 1983; Lowet et al., 2015; Lowet, Roberts, Peter, Gips, & De Weerd, 2017; Stettler et al., 2002; Ts'o et al., 1986). Rather than introducing a new explanatory variable, TWCO offers a mechanistic synthesis and shows how the established influence of these features on perception can emerge from the dynamics of coupled neural oscillators in V1. Thus, the success of these manipulations may arise precisely because they tap into the factors that determine whether synchrony can form among neural assemblies. As such, TWCO may provide a unifying principle that explains why these stimulus features are effective in modulating the efficiency of figure-ground segregation.

Future work should explore to what extent the principles of TWCO can explain segmentation of objects in natural images. While synchrony-based grouping mechanisms based on these principles have been used to segment natural images in machine vision (Fang et al., 2014; Lowet et al., 2015; Nikonov et al., 2020), it remains an open question whether cortical synchrony mediates human perception for such stimuli. Similarly, it remains an open question whether the principles outlined here generalize beyond the visual system to other sensory modalities. Interestingly, related forms of stimulus-dependent synchrony have been observed in auditory cortex, where it facilitates the integration of sound features and the segregation of auditory streams (Giraud & Poeppel, 2012). Finally, the principles of TWCO might provide a novel lens through which we can understand perceptual symptoms in neurological and psychiatric disorders. For example, schizophrenia is characterized by disrupted perceptual grouping and figure-ground segregation (Liddle, 1987; Malaspina et al., 2004; Uhlhaas et al., 2006) and disrupted visual gamma (Spencer et al., 2003). The prominent role of coupling strength within TWCO raises the possibility that reduced dendritic spine density in layer 3 of V1 within schizophrenia patients (Fish et al., 2025) may contribute to disrupted gamma synchrony and that this, in turn, may lead to disrupted perceptual grouping.

In conclusion, this study shows that figure-ground segregation performance can be well predicted by the factors that determine synchrony according to the theory of weakly coupled oscillators. Frequency detuning driven by contrast heterogeneity and coupling strength driven by physical distance may interact constructively to give rise to the perceptual skill of figure-ground segregation as well as its practice-induced enhancement. Our results show that a synchrony-based neural grouping mechanism can account for the observed behavioral patterns in a texture segregation task, and therefore remains a viable explanation for figure-ground segregation that cannot be ruled out. The documented dependence of gamma synchrony on stimulus features and element distance are essential components rather than obstacles to such a mechanism. This research sheds additional light on the underlying mechanisms of visual perception and perceptual learning and suggests that gamma oscillations and synchrony may be involved in the training-induced enhancement of figure-ground segregation.

Methods

Behavioral Experiments

The study and its experimental procedures were approved by the local Ethical Committee of the Faculty of Psychology and Neuroscience (ERCPN).

Participants

Eight healthy volunteers (6 female, mean age $= 23.75$, standard deviation $= 6.453$) participated in this study. Our study employed a repeated-measures design with extensive sampling, collecting a large number of trials from each participant. Sample size was determined based on comparable studies investigating visual perception and perceptual learning in humans (Intoy et al., 2024; Lange et al., 2020; Tesileanu et al., 2020). All participants had normal or corrected-to-normal visual acuity. After receiving full information about all procedures and the right to withdraw participation at any time, participants gave their written informed consent. All participants were compensated monetarily for their time.

Stimuli

Each texture stimulus consisted of a full-screen irregular grid of non-overlapping Gabor annuli with a diameter $0.7°$, a spatial frequency of 5.7 cycles/degree and a mean luminance of $60.76\ \mathrm{Cd/m^2}$ placed on a grey ($60.76\ \mathrm{Cd/m^2}$) background. Annuli contrasts were uniformly sampled from the full contrast range $U[0,1]$, except for a rectangular figure region [$(9 \pm 0.7)° \times (5 \pm 0.4)°$] whose contrasts were drawn from a second uniform distribution $U\left[0.5 - \frac{\zeta}{2},\ 0.5 + \frac{\zeta}{2}\right]$ with range ζ whose values were $\{0.01, 0.2575, 0.505, 0.7525, 1\}$. The figure region thus exhibited limited contrast heterogeneity, except when $\zeta = 1$ which is identical to the background (maximum) contrast heterogeneity. The center of the figure region was placed at an eccentricity of $(7 \pm 1)°$. The polar angle of the figure was varied on each trial with the condition that it was always completely inside a single visual field quadrant. The coarseness of the grid was expressed as a factor ρ that scales the average center-to-center distance between any pair of neighboring annuli in the whole texture. The values of ρ were $\{1, 1.125, 1.250, 1.375, 1.5\}$. Each annulus was initially placed on a regular grid and subsequently slightly shifted in a random direction by a distance chosen from a uniform distribution that ranged from zero to half of the edge-to-edge distances of neighboring annuli. All combinations of ζ and ρ yield 25 unique stimulus conditions.

Tasks and Procedure

The experiment consisted of nine consecutive sessions (eight training and one transfer session) with a two-alternative forced choice design in which participants were required to indicate whether the rectangular figure was oriented horizontally or vertically by pressing the right and left arrow key, respectively. Responses were given with the middle and index fingers of the right hand. Each trial of the experiment started with the presentation of a fixation point (a small bright turquoise disk of $2° \times 2°$) for minimally $1000\ \mathrm{ms}$, during which accurate fixation was to be initiated (i.e., deviation $< 2°$ from fixation point) to trigger stimulus presentation. Participants were required to maintain fixation throughout presentation of the stimulus ($1000\ \mathrm{ms}$ or less in case that a participant lost fixation or provided a response). Participants received feedback after each trial in the form of color changes

(green correct; red incorrect) of the fixation point lasting for $500\ \mathrm{ms}$. Feedback was followed by a $600\ \mathrm{ms}$ inter-trial interval during which an isoluminant (grey) screen was shown. When a participant's gaze fell outside the fixation window during the fixation period preceding the stimulus, or during stimulus presentation, the trial was aborted. Aborted trials were repeated at a randomly chosen time during the experiment.

The 25 conditions defined by contrast heterogeneity and grid coarseness were aggregated into experimental blocks such that all 25 combinations were shown exactly once per 25-trial block in random order. Each participant completed 30 blocks (750 trials) in each of the sessions. The figure was placed in the lower right quadrant for the eight training sessions. In the transfer session, the figure was moved to the orthogonal (upper left) quadrant. Participants were made aware of the figure displacement but were not told in which quadrant to expect it.

The experiment was conducted in a dimly lit room. A chin and headrest were used to support the participant's head and to keep eye-screen distance constant at $57\ \mathrm{cm}$. Stimuli were displayed on a $19''$ Samsung SyncMaster 940BF LCD monitor (Samsung, Seoul, South Korea; $60\ \mathrm{Hz}$ refresh rate, 1280×1024 resolution). Stimulus representation and response recording were performed by Psychtoolbox-3 for MATLAB 64-Bit (Version 3.0.14 - Build date: Apr 6th, 2018), under Microsoft Windows. Fixation was monitored with a desktop-mounted Eyelink 1000 eye-tracker (SR Research Ltd., $500\ \mathrm{Hz}$ or $1000\ \mathrm{Hz}$ sampling frequency, $< 0.01°$ RMS spatial resolution, eye-movement data were down-sampled to $250\ \mathrm{Hz}$).

Statistical Analyses

We used Bayesian hierarchical regression to analyze main and interactions of variables of interest which include manipulated stimulus features (contrast heterogeneity, grid coarseness and their interaction), model synchrony and learning effects (session and its interactions with stimulus features). Stimulus features and model synchrony were z-scored and session was mean-centered. Each of these statistical models included subjects-specific intercepts and slopes for all predictors. To investigate the relationship between the sizes of empirical and model Arnold tongues, we used Bayesian hierarchical regression with subject-level random intercepts. Because tongue size data entail only one measurement per subject per session, there was insufficient information to estimate subject-level slopes reliably.

All analyses were conducted using Bambi (v0.15.0) and ArviZ (v0.22.0) in Python. Priors were weakly informative defaults. Specifically, fixed effects were given Normal(0, 2.5), intercepts Normal(0, 2.5), and group-level standard deviations HalfNormal(2.5) priors. Correlations among random slopes and intercepts were given an LKJ(1) prior. Models were estimated using the No-U-Turn Sampler (NUTS) with 4 chains of 2,000 draws each, following a 2,000-draw tuning phase, for a total of 8,000 posterior samples. For all Bayesian models, convergence was assessed using standard diagnostics. All $\hat{R}$ values were approximately 1.00, and all effective sample sizes (ESS) were sufficient, with the smallest ESS across analyses being 2200.

Oscillator Model of V1

We model a small patch of V1 that receives input from a $6.7° \times 6.7°$ square region of the visual field. The area of this square region matches the area of the rectangular figure region in our psychophysics experiments. The center of this region is furthermore located at an eccentricity matching that of the figure. We model this V1 patch as a network of weakly coupled phase oscillators arranged on an $n \times n$ $(n = 20)$ irregular grid on the cortical surface. To that end we first defined a regular grid of

receptive field centers for each oscillator in visual space and subsequently transformed receptive field coordinates to cortical coordinates of V1 using a complex-l(Balasubramanian & Schwartz, 2002; Schwartz, 1980)Schwartz, 2002; Schwartz, 1980) with generic human parameter values ($a = 0.7, \alpha = 0.9$; Polimeni et al., 2005). While receptive fields are thus equally spaced in the visual field, neural oscillators themselves are not equally spaced on the cortical surface. The phase of each neural oscillator evolves according to a Kuramoto model:

$$\dot{\theta}_i = \omega_i + \frac{1}{N}\sum_{j=1}^{N} K_{ij}^{s} \sin\left(\theta_j - \theta_i\right), \quad i = 1, \dots, N; s = 1, \dots, 9 \tag{1}$$

where, θ_i is the phase of the ith oscillator, ω_i its intrinsic frequency, K_{ij}^{s} the coupling strength between oscillators i and j in session s (note that s is an index and not an exponent) and $N = n^2$ is the total number of oscillators. We treat the instantaneous frequency $\frac{\dot{\theta}}{2\pi}$ of each oscillator as a proxy for the instantaneous population firing rate of the corresponding neural assembly.

Intrinsic Frequency

In accordance with electrophysiological findings, the intrinsic frequency of each oscillator is a function of the local contrast in its receptive field (Roberts, Lowet, Brunet, TerWal, Tiesinga, Fries, & de Weerd, 2013). Specifically, the typical oscillation frequency ν (in Hz with corresponding $\omega = 2\pi\nu$) of a neural circuit in V1 is a linear function of local contrast (Lowet et al., 2015):

$$\nu = 25 + 0.25C \tag{2}$$

The local contrast received by each oscillator i is given by the weighted root-mean-squared (RMS) value of contrast (Frazor & Geisler, 2006):

$$C_i = \sqrt{\sum_{h=1}^{M} w_{ih} \frac{(L_h - \bar{L})^2}{\bar{L}^2}) \Big/ \sum_{h=1}^{M} w_{ih}} \tag{3}$$

where L_h is the luminance of pixel h in the stimulus, $\bar{L}$ is the mean luminance over all pixels, and w_{ih} is the weight of pixel h and oscillator i. The weighting was specific to each oscillator as it reflects its unique receptive field which we modelled using an isotropic 2D Gaussian function:

$$w_{ih} = exp\left(-\frac{(x_h - X_i)^2 + \left(y_h - Y_i\right)^2}{2{\sigma_i}^2}\right), \tag{4}$$

Here, (x_h, y_h) are the coordinates of the hth pixel, while (X_i, Y_i) are the coordinates of the receptive field center of the ith oscillator. In addition, σ_i is the size of the receptive field. We estimated receptive field sizes based on their location relative to the center of gaze. Specifically, receptive field diameter in V1 exhibits a threshold linear relationship with receptive field eccentricity (e; Freeman & Simoncelli, 2011) such that $\varnothing = \max(0.172\mathrm{e} - 0.25, 1)$. We related the receptive field diameter to the standard deviation of a Gaussian in two steps. First, we related the diameter to the full width at half maximum (FWHM) of a Gaussian beam $FWHM = \frac{\sqrt{ln2}}{\sqrt{2}}\varnothing$ (Hill, 2007). Then, we related the FWHM to the standard deviation $\sigma = \frac{FWHM}{2\sqrt{2ln2}}$. Combining these steps, the standard deviation is one fourth of the receptive field diameter.

Adaptive Coupling

The coupling strength K_{ij}^1 between pairs of oscillators in the first session is a function of their cortical distance:

$$K_{ij}^1 = \gamma e^{-\lambda d_{ij}}. \tag{5}$$

Here, γ is the maximum coupling strength and λ controls how fast the coupling strength decreases as a function of cortical distance (d_{ij}) between oscillators i and j. We estimated γ and λ from previously published data relating coupling strength to cortical distance within V1 in two macaque monkeys (Lowet, Roberts, Peter, Gips, & de Weerd, 2017).

Coupling strength in the remaining sessions is the result of an offline learning process that takes the experience accumulated over an individual training session into account. Specifically, learning in our model depends on the pairwise phase-locking value (PLV; Lachaux et al., 1999) between model oscillators accumulated over trials within one session. Phase-locking values were computed over the second half of the simulation period, which was subsampled to 50 timepoints. Accumulation across trials involves summing PLVs over trials, where the contribution of each trial is weighted by the probability that the model would produce a correct response on that trial. The weighted PLV is summarized in a matrix Q. To obtain the probability of a correct response (P_c) from model simulations, we related it to the degree of synchrony (r) among phase oscillators through a psychometric function:

$$P_c = 1/[1 + \exp(-\mu_0 - \mu_1 r)] \tag{6}$$

Parameters of this function (i.e., μ_0 and μ_1) were estimated based on model simulations and empirical results from the first session. The temporal evolution of pairwise coupling strength is given by a Hebbian-type learning rule:

$$\dot{K}_{ij} = \epsilon\left(\gamma Q_{ij}^s - K_{ij}\right). \tag{7}$$

Here, ϵ is a learning rate. Essentially, pairwise structural coupling approaches pairwise functional coupling, as measured by the weighted PLV within a session (Q^s), scaled by the maximum coupling strength γ. Integration of Equation 5 with respect to time yields

$$K_{ij}^{s+1} = exp(-\epsilon t)K_{ij}^{s} + [1 - exp(-\epsilon t)]\,\gamma Q_{ij}^{s}. \tag{8}$$

Here, t is the time between two sessions during which learning occurs (e.g., during sleep). Since neither t nor ϵ can be measured independently and are not known a priori, we merged them into a single free parameter $E = \epsilon t$. We refer to this as the effective learning rate. We adjusted the parameter E to maximize the correspondence, measured by the weighted Jaccard similarity, between the distribution of performance observed in the *second* experimental session and the distribution of synchrony after letting the model learn according to Equation 6. To that end we used a coarse-to-fine grid search wherein we let the model learn using a grid of 25 candidate effective learning rates and selected the value that enabled best prediction of session 2 performance. We then created a new, finer, grid around the best effective learning rate and repeated this procedure. In total, we explored five nested grids. Note that the learning procedure depends on data from the first two sessions to establish a mapping from synchrony to performance (parameters μ_0 and $\mu_1$of the psychometric function linking synchrony to performance) and to estimate the effective learning rate, respectively. We kept these parameters fixed for predicting the results of sessions 3-8. To further disentangle data used for parameter tuning and data used for testing model predictions, we utilized a leave-one-out cross-validation procedure. We estimated all parameters from the first two sessions of seven of our eight participants, and then predicted results of session 3-8 in the left-out participant. We repeated this procedure eight times, once per participant, and stored all results for further analysis.

Simulations

We simulated eight training sessions, each consisting of 30 blocks with 25 trials. Within each trial, we simulated a one-second stimulus monitoring interval assigned to a specific combination of contrast heterogeneity and grid coarseness. All simulations were performed in Python 3.12.2 using the odeint method from scipy's (version 1.12.0) integrate submodule. For each simulated trial, we evaluated synchrony by measuring the radius ($r \in [0,1]$) of the Kuramoto order parameter given by

$$re^{i\psi} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j} \tag{9}$$

where θ_j is the phase of the oscillator j. For each simulated trial, r was averaged within the second half of the trial duration, and over all blocks.

System Specifications

All analyses and simulations were performed as a Docker containerized Snakemake workflow executed on a single compute node of Maastricht University's Data Science Research Infrastructure (DSRI). The node is equipped with two AMD EPYC 7551 32-Core Processors, has a nominal 512 GB of RAM, and operates on Fedora 37. The workflow utilized 30 of 64 available cores to simulate all blocks of a particular trial in parallel. To ensure that all results can be reproduced exactly, the random seed of our workflow was fixed at 1709026616.

Data Availability

The data acquired in the study can be accessed at https://zenodo.org/doi/10.5281/zenodo.10817186.

Code Availability

The code for data acquisition can be accessed at https://github.com/ccnmaastricht/TextureStimuli-FigureGround.git.The code for performing all analyses and simulations can be accessed at https://github.com/ccnmaastricht/NeuralSynchrony-FigureGround.

References

Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F., & Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, *77*(1), 137–185. https://doi.org/10.1103/RevModPhys.77.137

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. https://doi.org/10.1038/387401a0

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Amir, Y., Harel, M., & Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *Journal of Comparative Neurology*, *334*(1), 19–46. https://doi.org/10.1002/cne.903340103

Anand, S., Cho, H., Adamek, M., Burton, H., Moran, D., Leuthardt, E., & Brunner, P. (2023). High gamma coherence between task-responsive sensory-motor cortical regions in a motor reaction-time task. *Journal of Neurophysiology*, *130*(3), 628–639. https://doi.org/10.1152/jn.00172.2023

Balasubramanian, M., & Schwartz, E. L. (2002). The isomap algorithm and topological stability. *Science*, *295*(5552), 7. https://doi.org/10.1126/science.1066234

Baldi, P., & Meir, R. (1990). Computing with Arrays of Coupled Oscillators: An Application to Preattentive Texture Discrimination. *Neural Computation*, *2*(4), 458–471.

Boucsein, C., Nawrot, M., Schnepel, P., & Aertsen, A. (2011). Beyond the Cortical Column: Abundance and Physiology of Horizontal Connections Imply a Strong Role for Inputs from the Surround. *Frontiers in Neuroscience*, *5*, 32. https://doi.org/10.3389/fnins.2011.00032

Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *Journal of Neuroscience*, *22*(5), 1976–1984. https://doi.org/10.1523/jneurosci.22-05-01976.2002

Brosch, T., Neumann, H., & Roelfsema, P. R. (2015). Reinforcement Learning of Linking and Tracing Contours in Recurrent Neural Networks. *PLOS Computational Biology*, *11*(10), e1004489. https://doi.org/10.1371/journal.pcbi.1004489

Buia, C., & Tiesinga, P. (2006). Attentional modulation of firing rate and synchrony in a model cortical network. *Journal of Computational Neuroscience*, *20*(3), 247–264. https://doi.org/10.1007/s10827-006-7074-2

Burkhalter, A., & Bernardo, K. L. (1989). Organization of corticocortical connections in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(3), 1071–1075. https://doi.org/10.1073/pnas.86.3.1071

Buzsáki, G., Logothetis, N., & Singer, W. (2013). Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms. *Neuron*, *80*(3), 751–764. https://doi.org/10.1016/J.NEURON.2013.10.002

Chen, X., Sanayei, M., & Thiele, A. (2013). Perceptual learning of contrast discrimination in macaca mulatta. *Journal of Vision*, *13*(13), 22–22. https://doi.org/10.1167/13.13.22

Coombes, S., & Bressloff, P. C. (1999). Mode locking and Arnold tongues in integrate-and-fire neural oscillators. *Physical Review E*, *60*(2), 2086–2096. https://doi.org/10.1103/PhysRevE.60.2086

Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: experience and attention in primary visual cortex. *Nature Neuroscience*, *4*(5), 519–525. https://doi.org/10.1038/87470

de Weerd, P., Sprague, J. M., Vandenbussche, E., & Orban, G. A. (1994). Two stages in visual texture segregation: a lesion study in the cat. *Journal of Neuroscience*, *14*(3), 929–948. https://doi.org/10.1523/JNEUROSCI.14-03-00929.1994

de Weerd, P., Vandenbussche, E., & Orban, G. A. (1992). Texture segregation in the cat: A parametric study. *Vision Research*, *32*(2), 305–322. https://doi.org/10.1016/0042-6989(92)90216-M

Doelling, K. B., & Florencia Assaneo, M. (2021). Neural oscillations are a start toward understanding brain activity rather than the end. *PLoS Biology*, *19*(5), e3001234. https://doi.org/10.1371/journal.pbio.3001234

Dubey, A., & Ray, S. (2020). Comparison of tuning properties of gamma and high-gamma power in local field potential (LFP) versus electrocorticogram (ECoG) in visual cortex. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-68857-6

Duecker, K., Gutteling, T. P., Herrmann, C. S., & Jensen, O. (2021). No Evidence for Entrainment: Endogenous Gamma Oscillations and Rhythmic Flicker Responses Coexist in Visual Cortex. *The Journal of Neuroscience*, *41*(31), 6684–6698. https://doi.org/10.1523/JNEUROSCI.3134-20.2021

Eckhorn, R. (1994). Oscillatory and non-oscillatory synchronizations in the visual cortex and their possible roles in associations of visual features. *Progress in Brain Research*, *102*, 405–426. https://doi.org/10.1016/S0079-6123(08)60537-8

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 14–14. https://doi.org/10.1167/11.5.14

Ermentrout, B., Park, Y., & Wilson, D. (2019). Recent advances in coupled oscillator theory. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *377*(2160). https://doi.org/10.1098/RSTA.2019.0092

Evers, K., Peters, J., & Senden, M. (2021). Cortical Synchrony as a Mechanism of Collinear Facilitation and Suppression in Early Visual Cortex. *Frontiers in Systems Neuroscience*, *15*, 73. https://doi.org/10.3389/fnsys.2021.661161

Fang, Y., Cotter, M. J., Chiarulli, D. M., & Levitan, S. P. (2014). Image segmentation using frequency locking of coupled oscillators. *International Workshop on Cellular Neural Networks and Their Applications*. https://doi.org/10.1109/CNNA.2014.6888657

Fernandez-Ruiz, A., Sirota, A., Lopes-dos-Santos, V., & Dupret, D. (2023). Over and above frequency: Gamma oscillations as units of neural circuit operations. *Neuron*, *111*(7), 936–953. https://doi.org/10.1016/j.neuron.2023.02.026

Fish, K. N., Sweet, R. A., MacDonald, M. L., & Lewis, D. A. (2025). Regional Specificity of Cortical Layer 3 Dendritic Spine Deficits in Schizophrenia. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2025.2221

Frazor, R. A., & Geisler, W. S. (2006). Local luminance and contrast in natural images. *Vision Research*, *46*(10), 1585–1598. https://doi.org/10.1016/J.VISRES.2005.06.038

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

Fries, P. (2009). Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annual Review of Neuroscience*, *32*(1), 209–224. https://doi.org/10.1146/annurev.neuro.051508.135603

Gilbert, C. D., & Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, *3*(5), 1116–1133. https://doi.org/10.1523/jneurosci.03-05-01116.1983

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, *9*(7), 2422–2432. https://doi.org/10.1523/JNEUROSCI.09-07-02432.1989

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Grossberg, S., & Mingolla, E. (1985). Neural Dynamics of Form Perception: Boundary Completion, Illusory Figures, and Neon Color Spreading. *Psychological Review*, *92*(2), 173–211. https://doi.org/10.1037/0033-295X.92.2.173

Grossberg, S., & Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading. *Computer Vision, Graphics, and Image Processing*, *37*(1), 116–165. https://doi.org/10.1016/S0734-189X(87)80012-0

Hadjipapas, A., Lowet, E., Roberts, M. J., Peter, A., & de Weerd, P. (2015). Parametric variation of gamma frequency and power with luminance contrast: A comparative study of human MEG and monkey LFP and spike responses. *NeuroImage*, *112*, 327–340. https://doi.org/10.1016/j.neuroimage.2015.03.007

Hall, S. D., Holliday, I. E., Hillebrand, A., Singh, K. D., Furlong, P. L., Hadjipapas, A., & Barnes, G. R. (2005). The missing link: Analogous human and primate cortical gamma oscillations. *NeuroImage*, *26*(1), 13–17. https://doi.org/10.1016/j.neuroimage.2005.01.036

Henrie, J. A., & Shapley, R. (2005). LFP power spectra in V1 cortex: The graded effect of stimulus contrast. *Journal of Neurophysiology*, *94*(1), 479–490. https://doi.org/10.1152/jn.00919.2004

Henriksson, L., Nurminen, L., Hyvärinen, A., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision*, *8*(10), 5. https://doi.org/10.1167/8.10.5

Hill, D. (2007). How to convert FWHM measurements to 1/e-squared halfwidths. *Radiant Zemax Knowledge Base*.

Hochstein, S., & Ahissar, M. (2002). View from the Top: Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Hua, T., Bao, P., Huang, C.-B., Wang, Z., Xu, J., Zhou, Y., & Lu, Z.-L. (2010). Perceptual Learning Improves Contrast Sensitivity of V1 Neurons in Cats. *Current Biology*, *20*(10), 887–894. https://doi.org/10.1016/j.cub.2010.03.066

Huang, L., Wang, L., Shen, W., Li, M., Wang, S., Wang, X., & Zhang, X. (2020). A source for awareness-dependent figure–ground segregation in human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(48), 30836–30847. https://doi.org/10.1073/pnas.2009232117

Kandel, E., Schwartz, J., Jessell, T., & Siegelbaum, S. (2000). *Principles of Neural Science* (5th ed., Vol. 4). McGraw-Hill.

Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, *7*(4), 530–535. https://doi.org/10.1016/S0959-4388(97)80033-5

Keil, M. S., Cristóbal, G., Hansen, T., & Neumann, H. (2005). Recovering real-world images from single-scale boundaries with a novel filling-in architecture. *Neural Networks*, *18*(10), 1319–1331. https://doi.org/10.1016/j.neunet.2005.08.002

Kirchberger, L., Mukherjee, S., Schnabel, U. H., van Beest, E. H., Barsegyan, A., Levelt, C. N., & Roelfsema, P. R. (2021). The essential role of recurrent processing for figure-ground perception in mice. *Science Advances*, *7*(27). https://doi.org/10.1126/sciadv.abf2701

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. *19*. https://doi.org/10.1007/978-3-642-69689-3

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). *Measuring phase synchrony in brain signals*. Human Brain Mapping. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*(2), 1605–1615. https://doi.org/10.1523/jneurosci.15-02-01605.1995

Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*(4), 529–535. https://doi.org/10.1016/S0959-4388(98)80042-1

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, *31*(4), 679–691. https://doi.org/10.1016/0042-6989(91)90072-H

Layton, O. W., Mingolla, E., & Yazdanbakhsh, A. (2014). Neural dynamics of feedforward and feedback processing in figure-ground segregation. *Frontiers in Psychology*, *5*, 972. https://doi.org/10.3389/fpsyg.2014.00972

Liddle, P. F. (1987). The Symptoms of Chronic Schizophrenia. *British Journal of Psychiatry*, *151*(2), 145–151. https://doi.org/10.1192/bjp.151.2.145

Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research*, *40*(1), 97–109. https://doi.org/10.1016/S0042-6989(99)00153-5

Lowet, E., Roberts, M., Hadjipapas, A., Peter, A., Eerden, J. van der, & Weerd, P. De. (2015). Input-Dependent Frequency Modulation of Cortical Gamma Oscillations Shapes Spatial Synchronization and Enables Phase Coding. *PLOS Comput Biol*, *11*(2), e1004072. https://doi.org/10.1371/journal.pcbi.1004072

Lowet, E., Roberts, M. J., Peter, A., Gips, B., & De Weerd, P. (2017). A quantitative theory of gamma synchronization in macaque V1. *ELife*, *6*. https://doi.org/10.7554/eLife.26642

Lund, J. S., Yoshioka, T., & Levitt, J. B. (1993). Comparison of Intrinsic Connectivity in Different Areas of Macaque Monkey Cerebral Cortex. *Cerebral Cortex*, *3*(2), 148–162. https://doi.org/10.1093/cercor/3.2.148

Malagon, G., Miki, T., Tran, V., Gomez, L., & Marty, A. (2020). Incomplete vesicular docking limits synaptic strength under high release probability conditions. *ELife*, *9*. https://doi.org/10.7554/eLife.51404

Malaspina, D., Simon, N., Goetz, R. R., Corcoran, C., Coleman, E., Printz, D., Mujica-Parodi, L., & Wolitzky, R. (2004). The Reliability and Clinical Correlates of Figure-Ground Perception in Schizophrenia. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *16*(3), 277–283. https://doi.org/10.1176/jnp.16.3.277

Masquelier, T., Hugues, E., Deco, G., & Thorpe, S. J. (2009). Oscillations, Phase-of-Firing Coding, and Spike Timing-Dependent Plasticity: An Efficient Learning Scheme. *Journal of Neuroscience*, *29*(43), 13484–13493. https://doi.org/10.1523/JNEUROSCI.2207-09.2009

Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., & Rodriguez, E. (2007). Synchronization of Neural Activity across Cortical Areas Correlates with Conscious Perception. *The Journal of Neuroscience*, *27*(11), 2858–2865. https://doi.org/10.1523/JNEUROSCI.4623-06.2007

Motoyoshi, I. (1999). Texture filling-in and texture segregation revealed by transient masking. *Vision Research*, *39*(7), 1285–1291. https://doi.org/10.1016/S0042-6989(98)00254-1

Motoyoshi, I., & Nishida, S. (2001). Temporal resolution of orientation-based texture segregation. *Vision Research*, *41*, 2089–2105. https://doi.org/10.1016/S0042-6989(01)00123-3

Neisser, U. (1964). VISUAL SEARCH. *Scientific American*, *210*(6), 94–103. https://www.jstor.org/stable/24931530?seq=1

Neu, J. C. (1979). Coupled Chemical Oscillators. *SIAM Journal on Applied Mathematics*, *37*(2), 307–315. https://doi.org/10.1137/0137022

Neumann, H., Pessoa, L., & Hansen, T. (2001). Visual filling-in for computing perceptual surface properties. *Biological Cybernetics*, *85*(5), 355–369. https://doi.org/10.1007/s004220100263

Nikonov, D. E., Kurahashi, P., Ayers, J. S., Li, H., Kamgaing, T., Dogiamis, G. C., Lee, H.-J., Fan, Y., & Young, I. A. (2020). Convolution Inference via Synchronization of a Coupled CMOS Oscillator Array. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, *6*(2), 170–176. https://doi.org/10.1109/JXCDC.2020.3046143

Nothdurft, H. C. (1985a). Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vision Research*, *25*(4), 551–560. https://doi.org/10.1016/0042-6989(85)90067-1

Nothdurft, H. C. (1985b). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, *25*(12), 1957–1968. https://doi.org/10.1016/0042-6989(85)90080-4

Nothdurft, H. C. (1991a). Different effects from spatial frequency masking in texture segregation and texton detection tasks. *Vision Research*, *31*(2), 299–320. https://doi.org/10.1016/0042-6989(91)90092-I

Nothdurft, H. C. (1991b). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*(6), 1073–1078. https://doi.org/10.1016/0042-6989(91)90019-K

Pessoa, L., & de Weerd, P. (2003). *Filling-In: From Perceptual Completion to Cortical Reorganization*. Oxford University Press.

Pikovsky, A., Rosenblum, M., Kurths, J., & Synchronization, A. (2001). A universal concept in nonlinear sciences. *Self*, *2*, 3.

Polimeni, J. R., Hinds, O. P., Balasubramanian, M., van der Kouwe, A. J. W., Wald, L. L., Dale, A. M., & Schwartz, E. L. (2005). Two-dimensional mathematical structure of the human visuotopic map complex in V1, V2, and V3 measured via fMRI at 3 and 7 Tesla. *Journal of Vision*, *5*(8), 898. https://doi.org/10.1167/5.8.898

Poort, J., Self, M. W., van Vugt, B., Malkki, H., & Roelfsema, P. R. (2016). Texture Segregation Causes Early Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex. *Cerebral Cortex*, *26*(10), 3964–3976. https://doi.org/10.1093/cercor/bhv180

Raiguel, S., Vogels, R., Mysore, S. G., & Orban, G. A. (2006). Learning to See the Difference Specifically Alters the Most Informative V4 Neurons. *The Journal of Neuroscience*, *26*(24), 6589–6602. https://doi.org/10.1523/jneurosci.0457-06.2006

Ray, S., & Maunsell, J. H. R. (2010). Differences in Gamma Frequencies across Visual Cortex Restrict Their Possible Use in Computation. *Neuron*, *67*(885). https://doi.org/10.1016/j.neuron.2010.08.029

Ray, S., & Maunsell, J. H. R. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends in Cognitive Sciences*, *19*(2), 78–85. https://doi.org/10.1016/J.TICS.2014.12.002

Roberts, M. J., Lowet, E., Brunet, N. M., TerWal, M., Tiesinga, P., Fries, P., & de Weerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron*. https://doi.org/10.1016/j.neuron.2013.03.003

Roberts, M. J., Lowet, E., Brunet, N. M., TerWal, M., Tiesinga, P., Fries, P., & DeWeerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron*, *78*(3), 523–536. https://doi.org/10.1016/j.neuron.2013.03.003

Roelfsema, P. R. (2023). Solving the binding problem: Assemblies form when neurons enhance their firing rate—they don't need to oscillate or synchronize. *Neuron*, *111*(7), 1003–1019. https://doi.org/10.1016/j.neuron.2023.03.016

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience*, *7*(9), 982–991. https://doi.org/10.1038/nn1304

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure - Ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience*, *14*(4), 525–537. https://doi.org/10.1162/08989290260045835

Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology*, *7*(7), 461–467. https://doi.org/10.1016/S0960-9822(06)00221-2

Sanayei, M., Chen, X., Chicharro, D., Distler, C., Panzeri, S., & Thiele, A. (2018). Perceptual learning of fine contrast discrimination changes neuronal tuning and population coding in macaque V4. *Nature Communications*, *9*(1), 4238. https://doi.org/10.1038/s41467-018-06698-w

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature 2001 412:6846*, *412*(6846), 549–553. https://doi.org/10.1038/35087601

Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research*, *20*(8), 645–669. http://www.sciencedirect.com/science/article/pii/0042698980900905

Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology*, *17*(2), 148–153. https://doi.org/10.1016/J.CONB.2007.02.004

Self, M. W., Kooijmans, R. N., Supèr, H., Lamme, V. A., & Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(27), 11031–11036. https://doi.org/10.1073/pnas.1208097109

Shapira, A., Sterkin, A., Fried, M., Yehezkel, O., Zalevsky, Z., & Polat, U. (2017). Increased gamma band activity for lateral interactions in humans. *PLoS ONE*, *12*(12), e0187520. https://doi.org/10.1371/journal.pone.0187520

Shapley, R., & Hawken, M. J. (2011). Color in the Cortex: Single- and double-opponent cells. In *Vision Research* (Vol. 51, Issue 7, pp. 701–717). Pergamon. https://doi.org/10.1016/j.visres.2011.02.012

Shirhatti, V., Ravishankar, P., & Ray, S. (2022). Gamma oscillations in primate primary visual cortex are severely attenuated by small stimulus discontinuities. *PLOS Biology*, *20*(6), e3001666. https://doi.org/10.1371/journal.pbio.3001666

Singer, W. (1999). Neuronal Synchrony: A Versatile Code for the Definition of Relations? *Neuron*, *24*(1), 49–65. https://doi.org/10.1016/S0896-6273(00)80821-1

Spencer, K. M., Nestor, P. G., Niznikiewicz, M. A., Salisbury, D. F., Shenton, M. E., & McCarley, R. W. (2003). Abnormal Neural Synchrony in Schizophrenia. *The Journal of Neuroscience*, *23*(19), 7407–7411. https://doi.org/10.1523/JNEUROSCI.23-19-07407.2003

Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral Connectivity and Contextual Interactions in Macaque Primary Visual Cortex. *Neuron*, *36*(4), 739–750. https://doi.org/10.1016/S0896-6273(02)01029-2

Strogatz, S. H. (2000). From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, *143*(1–4), 1–20. https://doi.org/10.1016/S0167-2789(00)00094-4

Supèr, H., & Lamme, V. A. F. (2007). Altered figure-ground perception in monkeys with an extra-striate lesion. *Neuropsychologia*, *45*(14), 3329–3334. https://doi.org/10.1016/j.neuropsychologia.2007.06.015

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience*, *6*(4), 1160–1170. https://doi.org/10.1523/JNEUROSCI.06-04-01160.1986

Uhlhaas, P. J., Haenschel, C., Nikolić, D., & Singer, W. (2008). The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. In *Schizophrenia Bulletin* (Vol. 34, Issue 5, pp. 927–943). Oxford Academic. https://doi.org/10.1093/schbul/sbn062

Uhlhaas, P. J., Phillips, W. A., Mitchell, G., & Silverstein, S. M. (2006). Perceptual grouping in disorganized schizophrenia. *Psychiatry Research*, *145*(2–3), 105–117. https://doi.org/10.1016/j.psychres.2005.10.016

Voges, N., Schüz, A., Aertsen, A., & Rotter, S. (2010). A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Progress in Neurobiology*, *92*(3), 277–292. https://doi.org/10.1016/j.pneurobio.2010.04.001

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of Neuronal Interactions Through Neuronal Synchronization. *Science*, *316*(5831), 1609–1612. https://doi.org/10.1126/science.1139597

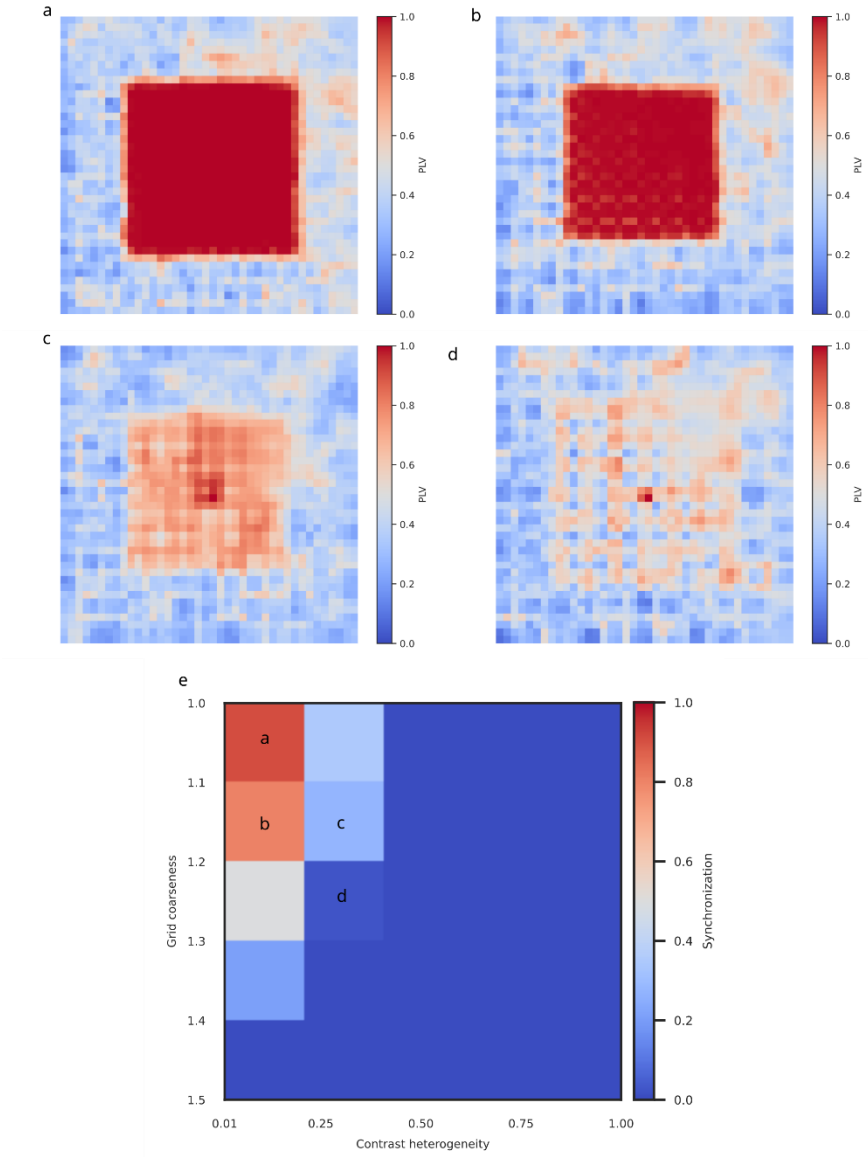
Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *The Journal of Neuroscience*, *24*(7), 1617–1626. https://doi.org/10.1523/JNEUROSCI.4442-03.2004

Zachariou, M., Roberts, M. J., Lowet, E., De Weerd, P., & Hadjipapas, A. (2021). Empirically constrained network models for contrast-dependent modulation of gamma rhythm in V1. *NeuroImage*, *229*, 117748. https://doi.org/10.1016/j.neuroimage.2021.117748

Supplementary Materials

Validation of the Figure-Only Model

To validate our figure-only modeling approach, we conducted simulations that included both the figure and a surrounding high-heterogeneity background. We visualized the emergence of a synchronized assembly by calculating the Phase-Locking Value (PLV) of every oscillator relative to a reference oscillator at the center of the figure.



Supplementary Figure 1: a-d, Phase-locking values of every oscillator relative to a reference oscillator positioned at the center of the figure for four different stimulus conditions. Phase-locking values are averages over 20 simulations. **e**, The Arnold Tongue from our figure-only simulations reference, with labels indicating the four representative conditions selected for the full simulation.

Design analysis

Table 2: Design analysis of main analysis in session 1. Detection probability refers to the proportion of simulated datasets in which the posterior probability of an effect exceeded 0.95 in the predicted direction. Type-S error indicates the probability of detecting an effect, but in the wrong direction (sign reversed). Type-M error refers to the ratio of estimated to true effect size when detected. A value of 1 indicates no deviation whereas values larger (smaller) than 1 indicate that effects are over (under) estimated.

Sample size	Detection probability			Type-S error			Type-M error		
	Contrast Heterogeneity	Grid Coarseness	Interaction	Contrast Heterogeneity	Grid Coarseness	Interaction	Contrast Heterogeneity	Grid Coarseness	Interaction
4	0.74	0.54	0.48	0.00	0.00	0.00	0.94	1.10	1.11
6	0.92	0.82	0.84	0.00	0.00	0.00	1.11	1.01	1.03
8	0.96	0.92	0.92	0.00	0.00	0.00	0.99	0.96	1.02

With a sample size of 8 participants, the detection probability (power) exceed 90% for all effects and robustness to both type-S and type-M errors.

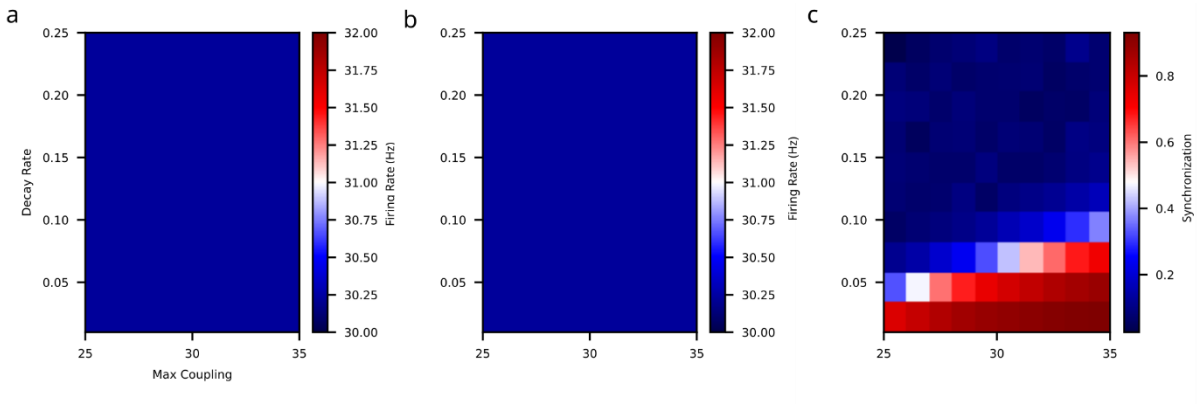
Average Firing is Insensitive to Oscillator Interactions

We demonstrate that average feedforward (intrinsic) firing rates and average effective firing rates emerging from oscillator interactions within our model are identical. We first show that average intrinsic and effective firing rates are identical across all 25 experimental conditions, using the fixed coupling parameters (γ=24.63, λ=0.22) used in the main text. For each of the 25 conditions, we computed the mean intrinsic and mean effective firing rates across all oscillators; i.e., across the entire figure region. The total absolute difference between average firing rates summed over all 25 conditions was negligible ($< 10^{-12}$ Hz).

To demonstrate that this is a general property of the model and not specific to our chosen parameter set, we performed a second analysis. We simulated a single, representative stimulus condition (contrast heterogeneity = 0.5, grid coarseness = 1.5) across a range of global coupling parameters (10 levels of maximum coupling strength γ and 10 levels of coupling decay λ).

For each of the resulting 100 simulations, we calculated the average intrinsic and effective firing rates across all oscillators. The average effective firing rate remained identical to the average intrinsic firing rate across this entire parameter space. The total absolute difference summed over all parameter combinations was again negligible ($< 10^{-11}$ Hz). This result is visualized in Figure S1. While synchrony varies across the parameter space, average firing rate remains constant and is determined exclusively by the stimulus.

Note that this does not imply that firing rates of individual oscillators are not affected by interactions with other oscillators. The fact that they can synchronize for certain stimulus conditions and/or parameter combinations shows that individual firing rates change, but they vary consistently around a fixed mean. Interactions in the model reduce the variance of firing rates (i.e., produce synchrony).



Supplementary Figure 2: Model derived quantities for different combinations of maximum coupling and decay rate. a, intrinsic firing rate averaged over all oscillators. **b**, effective firing rate averaged over all oscillators. **c**, In-phase synchronization among all oscillators. Neither average intrinsic nor average effective firing rates are sensitive to model parameters and are both fixed at 30.34 Hz.